1831 1958

THE ADVANCEMENT OF SCIENCE

VOL. XV NO. 58 SEPTEMBER 1958

Editorial Committee

Dr. H. HAMSHAW THOMAS, M.B.E., F.R.S. (*Chairman*)
Prof. L. DUDLEY STAMP, C.B.E., Dr. W. E. SWINTON, Prof. W. WARDLAW, C.B.E.

Sept. 1958 - Mar. 1960

R505 v. 15-16 (no. 58-64)

CONTENTS

Addresses delivered at the Annual Meeting of the British Association for the Advancement of Science in Glasgow, August 27 to September 3, 1958

PAGE

PORTRAITS OF THE PRESIDENT OF THE ASSOCIATION AND OF PRESIDENTS OF SECTIONS . . . . . *frontispiece*

OFFICERS AND COUNCIL 2

SCIENCE AND BUSINESS: A BALANCED PARTNERSHIP, *by* Sir Alexander Fleck, K.B.E., F.R.S. 3

SOME ASPECTS OF THERMONUCLEAR RESEARCH, *by* Dr. T. E. Allibone, F.R.S. 13

THE TRANSURANIC ELEMENTS, *by* Prof. H. J. Emeléus, C.B.E., F.R.S. . 23

BENEATH THE EARTH'S CRUST, *by* Prof. L. R. Wager, F.R.S. . . . 31

SOME BIOLOGICAL EFFECTS OF LOW TEMPERATURES, *by* Dr. A. S. Parkes, C.B.E., F.R.S. 46

THE ATLANTIC ENDS OF EUROPE, *by* Prof. E. Estyn Evans . . . 54

INFLATION AND THE BRITISH ECONOMY, *by* Prof. A. J. Brown . . . 65

HEAT AND THE ENGINEER, *by* Prof. O. A. Saunders, F.R.S. . . . 76

THE STUDY OF FOLK LIFE: AND ITS PART IN THE DEFENCE OF CIVILISATION, *by* Dr. Iorwerth C. Peate 86

SOME ASPECTS OF CELLULAR DIFFERENTIATION, *by* Dr. E. N. Willmer . 95

A REVIEW OF SOME PROBLEMS OF COLOUR VISION AND COLOUR BLINDNESS, *by* Prof. R. W. Pickford 104

THE CAPE AS AN ANCIENT AFRICAN FLORA, *by* Prof. R. S. Adamson . . 118

THE PARTNERSHIP BETWEEN INDUSTRY AND EDUCATION, *by* Sir Willis Jackson, F.R.S. 128

DIGESTION IN THE RUMINANT, *by* Dr. D. P. Cuthbertson, C.B.E. . . 140

JAN 3 1961

OFFICERS AND COUNCIL OF THE BRITISH ASSOCIATION, 1958

PATRON

HER MAJESTY THE QUEEN

PRESIDENT, 1958

SIR ALEXANDER FLECK, K.B.E., F.R.S.

GENERAL TREASURER

LORD RENNELL OF RODD, K.B.E., C.B.

GENERAL SECRETARIES

PROF. WILLIAM WARDLAW, C.B.E. DR. N. C. WRIGHT, C.B.

ORDINARY MEMBERS OF THE COUNCIL

J. N. L. BAKER
R. J. BARTLETT
Prof. L. F. BATES, F.R.S.
Dr. H. O. BULL
Prof. W. BURNS
P. RITCHIE CALDER, C.B.E.
Prof. G. R. CLEMO, F.R.S.
Prof. P. SARGANT FLORENCE
Prof. T. N. GEORGE
Sir RONALD GOULD
A. C. HARTLEY, C.B.E.
Prof. L. HUNTER
Rev. Prof. E. O. JAMES
Sir BEN LOCKSPEISER, K.C.B., F.R.S.
Dame KATHLEEN LONSDALE, D.B.E., F.R.S.
Prof. A. D. MACDONALD
JAMES MACDONALD, C.B.E.
L. R. MISSEN, C.M.G., M.C.
Dr. K. P. OAKLEY, F.B.A.
Dr. C. F. A. PANTIN, F.R.S.
Dr. R. W. POWELL
Dr. H. G. SANDERS
Dr. R. E. SLADE
Dr. W. E. SWINTON
Sir JAMES TURNER
Dr. W. B. TURRILL, O.B.E.
Prof. P. E. VERNON
W. H. WILCOCKSON
Dr. T. WILSON, O.B.E.
Prof. S. W. WOOLDRIDGE, C.B.E.

EX-OFFICIO MEMBERS OF THE COUNCIL

Past-Presidents of the Association, the President for the year, the President and Vice-Presidents for the ensuing Annual Meeting, past and present General Treasurers and General Secretaries, and the Local Treasurers and Local Secretaries for the Annual Meetings immediately past and ensuing.

PAST PRESIDENTS OF THE ASSOCIATION

H.R.H. THE DUKE OF WINDSOR, K.G., F.R.S. (1926)
Sir HENRY DALE, O.M., G.B.E., F.R.S. (1947)
Sir HENRY TIZARD, G.C.B., F.R.S. (1948)
Sir JOHN RUSSELL, O.B.E., F.R.S. (1949)
Sir HAROLD HARTLEY, G.C.V.O., C.B.E., M.C., F.R.S. (1950)
H.R.H. THE PRINCE PHILIP, DUKE OF EDINBURGH, K.G., K.T., C.B.E., F.R.S. (1951)
Prof. A. V. HILL, C.H., O.B.E., F.R.S. (1952)
Sir EDWARD APPLETON, G.B.E., K.C.B., F.R.S. (1953)
The Lord ADRIAN, O.M., F.R.S. (1954)
Sir ROBERT ROBINSON, O.M., F.R.S. (1955)
Sir RAYMOND PRIESTLEY, M.C. (1956)
Prof. P. M. S. BLACKETT, F.R.S. (1957)

PAST GENERAL OFFICERS OF THE ASSOCIATION

Prof. P. G. H. BOSWELL, O.B.E., F.R.S.
Prof. Sir DAVID BRUNT, Sec. R.S.
Sir HAROLD HARTLEY, G.C.V.O., C.B.E., M.C., F.R.S.
Dr. EDWARD HINDLE, F.R.S.
Sir FRANK SMITH, G.C.B., G.B.E., F.R.S.
Sir RICHARD SOUTHWELL, F.R.S.
Prof. F. J. M. STRATTON, O.B.E., F.R.S.

HON. AUDITORS

Dr. D. B. HARDEN, O.B.E. Dr. H. G. MAULE

SECRETARY

Sir GEORGE V. ALLEN, C.B.E.

SCIENCE AND BUSINESS: A BALANCED PARTNERSHIP[1]

ADDRESS BY SIR ALEXANDER FLECK, K.B.E., F.R.S.
PRESIDENT OF THE ASSOCIATION

My Lord Provost, My Lords, Ladies and Gentlemen,

I have to thank you for the welcome you have given to the British Association on our return to this City after the passage of some thirty years. The President on that last occasion was Sir William Bragg, and it is good to know that his son maintains a fine scientific tradition as Director of the Royal Institution and of the Davy Faraday Laboratory.

We are glad to have this opportunity in the course of our scientific deliberations to see something of the progress that has taken place in Glasgow and its region since then. In 1928 the Scottish Lowlands were reasonably prosperous: they had not yet entered the lean and hungry years of industrial depression when men's labours ground to a standstill and unemployment paralysed much initiative. As you well know, the depression lasted until early 1939, to be followed by the trials of war, death and destruction from the air, and all the while the deluding sense of prosperity that war brings. Since the end of the war, Glasgow has seen great industrial activity, though as we meet this evening, there are again threatening clouds in the sky.

But Glasgow's story goes back far beyond the fluctuating fortunes of the last thirty years. The little Cathedral town which emerged from the sixth-century settlement beside the Molendinar has become one of the greatest cities of the Commonwealth. The shallow, unpolluted Clyde has become the deep-dredged channel which has carried the world's largest ships, a source of trade and prosperity to Scotland.

We of the British Association pay tribute to the creative scientific thought of the citizens of Glasgow, which has enriched the world: first, Adam Smith, original economic thinker, who stood between the two storms that broke over Scotland—the Reformation and the Industrial Revolution; then the chemist, Joseph Black, whose work on the theory of heat helped to inspire James Watt, himself prime mover of the Industrial Revolution. Lister, one of our past Presidents, came to revolutionise surgical methods; Kelvin, another President, established principles of thermodynamics and electricity, and so 'put a girdle round about the earth'. In our own time, Soddy's work on atoms helped to name the new Atomic Age, presently so much in our minds.

These are some of the men who towered above their contemporaries and their century, men whose names make the British Association proud to renew its contact with this City and recall the debt that the world owes to them.

My Lord Provost, in thanking you for the warm welcome we have received from the City and the University, we express the hope that the citizens of Glasgow will always in the same inspiring tradition promote the advancement of science, technology and philosophic thought.

I am very sensible of the honour that the Council of the British Association have done me by electing me President for this Glasgow meeting and so enabling me to come back with them to the place in which I first saw the light of day just on the south bank of the Clyde. The surroundings there were all of trade and business, so I trust I can with appropriateness express some general thoughts on Business and Science and the need for a proper balance between them. This subject is very different from those chosen by most Presidents who have addressed this Association in

[1] Address delivered on August 27, 1958, at the Glasgow Meeting of the British Association.

the 127 years of its existence: they have spoken from the wealth of their expert knowledge of some science to which they have added great lustre. There is no science with which I, on the other hand, can claim the intimate acquaintance of a practising scientist. It is many years since I worked at a laboratory bench and experienced the excitement of anticipation all research workers feel as a well-planned experiment approaches its conclusion. That phase passed into one in which I supervised and directed others to do such work, and later that also was left behind under the pressures of full-time administration. The result of all this is that I am no specialist in any branch of science. Rather, I approach my task this evening as a sort of common man of science, who, lacking the detailed knowledge of the specialist, has at least an understanding of the framework within which scientific thought evolves.

All the same, some of you may feel that in my dual capacity as your President and as an administrator, I run the risk of finding myself a businessman among scientists and a scientist, even if only a lapsed scientist, among businessmen. Perhaps I may remind you of a remark ascribed to Talleyrand: 'War is much too serious a thing to be left to military men.' I have the impression that the same sentiment may apply both to the conduct of business and to the advancement of science. The risk of being neither pre-eminently scientist nor pre-eminently businessman may be counterbalanced by advantages that come from a non-specialist approach to science and business. I hope to show why such an approach may be particularly appropriate to the needs of the moment.

Science

In a population which depends so much on science, it is sad to think how few people clearly understand what it is all about. It is an approach to an understanding of the universe along a broad highway of organised knowledge firmly founded on observation and experiment embedded in a matrix of theory. It is a highway which is constantly being widened and extended and repaired. We are building it because we think that by and large it is a good road for mankind to travel, and that it is leading him away from ignorance and wretchedness towards a fuller life and a better world. Some find that the act of extending the highway gives them all they seek from their life-work: these are the pure scientists and much of our progress derives from them. Others find greater satisfaction in utilising what has already been built, pausing perhaps to strengthen it here, widen it there: these are the technologists or applied scientists. A few people combine both functions, driving the road forward into undiscovered territory, then paving it ready for all to use. Whatever our own work may be, we have a joint responsibility to see that the highway is properly and extensively used by all mankind. It is therefore fitting that the British Association now devoting itself to the additional task of securing the greater use of this highway by all types of people in these islands, more especially as we enter what is often called the Second Industrial Revolution.

In the first Industrial Revolution man learnt to convert chemical energy, derived from the reaction between carbon and oxygen, not merely into heat but into mechanical energy, steam being the key to the process. In the Second Industrial Revolution we are turning from the external properties of carbon atoms to the internal properties of the lightest and heaviest atoms in nature as a source of energy to expend as prodigally as we like, thus husbanding carbon atoms as a source of materials.

Another feature of the Second Industrial Revolution is the growing interdependence of the sciences as they advance on a broad front. To review the achievements of the last year or two alone, the launching of the Russian and American earth satellites depended on prior developments in several different sciences and technologies. Chemical rocket fuels, refractory combustion chambers, and radio transmitters which despatched back to the earth information of great relevance to the

theme of the International Geophysical Year, all played their part. Miniature transmitters themselves followed the invention in the Bell Telephone Laboratories ten years ago of the transistor, which in turn owes its successful development, at least in part, to the perfection of methods for separating traces of boron and other elements from silicon to a degree of purity exceeding one part in a thousand million.

All of us here recognise the creative possibilities of thermonuclear fusion energy opened up by the Harwell and Aldermaston experiments on high temperature electric discharges. Although most of the neutrons so far produced have not originated from thermonuclear fusion, there can be no doubt that ZETA and SCEPTRE are symbols of man's reach into a future of boundless energy; his grasp of such a prize could well rank as his greatest material achievement in the history of the world. In such experiments as these, as indeed with the discovery of nobelium, a prerequisite of success is a flourishing electrical industry which can provide many of the tools of nuclear physicists.

Scientific effort on this scale costs money; indeed, the total annual research expenditure in this country alone is estimated to be £300 million. Such effort also costs scientists. Public expenditure on education (including the Universities) is now some £740 million a year; if our Universities are to grow large enough to accommodate everyone who could benefit by such an education, instead of merely the 3½ per cent of the population who enjoy it now, then we shall need to contemplate spending something nearer £1,200 million a year. These are large sums and if as a nation we are to stay solvent they can only come from one source, the wealth which we produce.

BUSINESS

In this context it is easy to see the importance and responsibility of business, which produces and distributes most of our national wealth. Our capacity to extend our science and education largely depends on our ability to run business successfully. Lewis Mumford in his *Technics and Civilization*, written in the 1930's, though he appears to exaggerate their numerical importance, suggests that it may have been no accident that there were merchants of the City of London who played a part in the early days of the Royal Society. He draws attention to the affinity of science and business when he writes: 'The power that was science and the power that was money were, in final analysis, the same kind of power: the power of abstraction, measurement, quantification.'

On an occasion such as this it is difficult to avoid being at once too general and too particular. I have chosen to speak of business because business is a more comprehensive term than industry. At the same time I intend to limit my scope primarily to the larger units, although naturally I recognise the importance of the smaller organisations—indeed, in manufacturing industry alone, three-quarters of all establishments employ fewer than 100 people.

Defining the limits of my subject in this somewhat arbitrary way will for the moment do service for a formal definition of the term 'business'. After all, it is a vague word, perhaps for euphemistic reasons in the past, even deliberately vague. Nevertheless, I venture to put forward three characteristics which I think a business should have if it is to survive and function effectively over a period of years.

(*a*) First, it must contribute to the well-being of the community and its social evolution. I believe it can best do that in three specific ways: by producing goods or services for the community at fair prices; by giving fair treatment in the way of remuneration and working conditions to all who give their services to the business, whether as pay-roll employees, administrators or scientific staff; and by giving a fair return to those who risk their money in the business, so that, apart from this consideration, they will be encouraged to provide again for further ventures. When

business takes on itself the deployment of financial, material and human resources, it also takes on, I believe, the responsibility for seeing that they are properly and effectively used. In other words, it has an obligation to the community; and only in so far as it fulfils that obligation does it contribute to the well-being of the community and justify its own existence.

(*b*) Second, business must have the capacity to change its methods and if necessary even its objectives. It must be constantly alert to see that its products or services meet or anticipate the needs of the public, and to withstand the nipping and eager air of competition whether it comes from advancing techniques or from new entrants into the field.

(*c*) Finally, if it is to be self-perpetuating through growth or modernisation, a business must dispose of its products or services for more than their cost—or in plain language, it must make a profit. That is true of business whether run by private enterprise or by the State; nor does the political system of the State make any difference to the validity of this notion. Only from the excess of income over expenditure can we finance the expansion of business and science.

This brings me again to the reason why I have chosen to talk about business rather than industry. It is an old saying that money makes money; by the same token the production of wealth requires the investment of wealth, and such investment can come from a variety of sources. Traditionally part comes from retained profits, that is from what economists in the past called the abstinence of the individual manufacturer, or in today's terms, from the undistributed earnings which companies plough back into the business. Part comes from the banks, though their function in this country has usually been to provide working capital in the form of loans and overdrafts, rather than to provide fixed capital like the industrial banks of Germany which played such a prominent part in the organisation and growth of industry in that country. Indeed, we can share Disraeli's surprise that the London money market played such a small part in the financing of the railways. In *Endymion* published in 1880, he wrote: 'What is remarkable in this vast movement is that the great leaders of the financial world took no part in it . . . even Lombard Street which never was more wanted, was inactive, and it was only by the irresistible pressure of circumstances that a banking firm, which had an extensive country connection, was ultimately forced to take the leading part that was required.' Today by far the largest volume of capital comes from the investing public as private individuals or financial institutions, working with a vigorous and well-informed Stock Exchange.

There is, of course, no exact relationship between the amount of income which a country ploughs back into capital formation and the rate of economic growth which it achieves. Nevertheless, I think that certain generalisations can be offered. For example, we can compare the experience of the United States with that of this country. Between 1870 and the First World War, the United States devoted more than one-fifth of its National Product to home investment, and the United Kingdom only about one-tenth. During that period its National Product grew by more than 4 per cent per annum compared with less than 2½ per cent here. More recent history, in spite of occasional and, one hopes, temporary setbacks here, suggests the same lesson: over the past forty years the United States has invested about half as much again in relation to its resources as we have, and its National Product has risen at an average rate of about 3 per cent per annum, against only about 2 per cent in this country.

I do not wish to strike a note of disappointment so much as to remind you of the classic virtues of abstinence, saving and investment. In fact during the past ten years the amount which the United Kingdom has invested in its productive resources has shown an encouraging growth, with the apparent result that output in manufacturing industry, as a whole, has increased by some 3½ per cent per annum since 1948, compared with about 2½ per cent between the wars. If science has laid

the foundation for much of this expansion, business has provided the will and the means to complete the structure, and both must continue to do so.

There is another reason why it is important that this partnership should continue to flourish. We have in these islands over fifty million people who enjoy a material standard of living surpassed in few other countries in the world. Our mineral resources are limited and indeed declining. It is by means of the goods produced by manufacturing industry and the services provided by finance and commerce, in short by business as a whole, that our fifty millions are enabled to live, and on the whole to live tolerably well, in an area too small to feed them, at least by present methods.

All this is commonplace, but like much that is commonplace, it is easily taken for granted until we remember that over 90 per cent of the world's population live in countries less prosperous than our own, most of them in conditions that we should not dream of tolerating. The world does not owe us a living; we can only enjoy the standards which our business enterprise and scientific knowledge enable us to earn through our exports. Like all countries whose livelihood depends on their export trade, we have a vested interest in economically and politically stable conditions throughout the world; and I mean stable, not static, conditions. Stability will not be achieved for example, if plans for industrialisation of the less developed territories are obstructed from a misconceived notion that we may thereby be able to retain our own markets for that much longer. To those who heard Professor Blackett's plea last year for technological aid to the underdeveloped countries, this may sound an unnecessarily narrow approach. But I have deliberately couched my argument in terms of expediency rather than moral compulsion, because it emphasises one respect in which the combination of science and business seems to me to be vital, and it is this: we must concentrate for our own survival on the development of new products and new processes.

Our old staple exports of textiles, iron and steel, and coal, have long since given way to newer lines, vehicles, aircraft, machinery, chemicals and electrical goods. Even here the pattern keeps changing. Motor vehicles, fine and heavy chemicals and so on, are increasingly being manufactured in countries which previously produced only food or primary raw materials. In India, I am told, a company manufacturing Diesel engines and electric motors is exporting 30 per cent of its output. Chemicals from Red China and mass-produced toys and torches from Hong Kong are met in many of the world's markets. We are therefore obliged to rely either on selling things which no one else has yet learnt to make, or make as cheaply, or on marketing our skill and know-how.

It is not difficult to find examples of what happens to our business when 'Rule of Thumbism' is preferred to operations guided and controlled by science and technology. I think there would be general agreement that at various times some of our traditional industries in which we were pioneers, have suffered from the lack of proper scientific direction.

In my own industry, organic chemicals once experienced the same neglect. The lead that Perkin had given to our synthetic dyestuffs industry with mauveine, was frittered away through lack of scientific effort. In the Presidential Address to the Chemical Society in 1915 his son, William Henry Perkin, Junior, referred to this neglect in these words:

> My experience of the manufacturer in this country is that he is usually merely a commercial person who does not like the expert, and especially the idea of giving the expert a prominent position in the control of his works. Possibly the reason in many cases is ignorance of the value of science, but more probably it is due to the fact that being ignorant of science himself, he feels that if the expert is given too much prominence he must either study himself in order to understand the expert or leave the essential control of the business in his hands. Both these courses are distasteful to the ordinary commercial member of

a board of directors; the expert is therefore relegated to the background, and the business comes to grief.

The lesson has not yet been fully learned, for even today we meet this attitude. I have a quotation here from a Managing Director who went on record last May. The scope of his company may be small and may be provincial, but that the sentiment still exists in minds capable of directorial control is significant. He writes: 'The place of the expert is not in the saddle but as the assistant and the adviser. . . . The scientist must be on tap, not on top.' Such a relationship is very different from the partnership of science and business that I have in mind.

I have considered briefly some reasons why I regard the partnership of science and business as fundamental. To recapitulate, first it is because, if science is to carry forward our social evolution, it needs the financial and productive effort of business; and second, because the prosperity of this country depends on the technologically sophisticated products of our export trade.

A Balanced Partnership: How to Achieve It

(*a*) *The scientist in the business structure*

I should now like to examine how this partnership can be achieved in practice, by considering the place of the scientist in the business structure. In a technological business we can distinguish some six or seven separate functions, depending to some extent on their definition: finance, commerce, personnel, operation or production, design and construction, research and development. I propose not to elaborate on these functions fully, but to touch on one or two of their features.

First, research: research has been called the life-blood of industry, but I prefer to liken it to the pituitary gland which, I am told, after doing little that is obvious for several years except consume nourishment, at length stimulates other parts of the organism to develop and makes it capable of production.

Research applied to business can take different forms, and I think you may be interested in one example within my experience. It began with a research manager's conviction that a new inhalant anaesthetic was needed which should have the following properties: it should be a volatile liquid, with high anaesthetic potency, and be free from toxic, explosive and fire hazards. These criteria, he argued, pointed the way to aliphatic halogen compounds containing little hydrogen and especially to compounds with two or three fluorine atoms on the same carbon atom—groupings which are known by organic chemists to be not only unreactive in themselves, but also to strengthen neighbouring carbon-halogen bonds. The resulting inertness is one of the factors in avoiding toxicity; and the use of fluorine rather than other halogens keeps the volatility high. He also proposed a new method of expressing dosages which made it possible to relate, by using a principle of chemical thermodynamics, the amount of anaesthetic in the inspired air to that in the body fluids. This helped in the selection, for biological testing, of compounds likely to have the desired potency, among which the one finally chosen was 2-bromo-2-chloro-1:1:1-trifluoroethane, a new compound synthesized specially for the purpose.

The major research effort, which for the first two years fell to the organic and physical chemists, had at this stage passed to the pharmacologists, whose screening tests took some three years. They showed that the new compound had not only the properties sought, including a high margin of safety from overdose, but also some unexpected advantages. Clinical trials with successful results began six years after the research started, and meanwhile manufacturing methods were developed by chemists and engineers, leading to general release to the medical profession two years later.

There are several features about this much abbreviated story, some of which are relevant to my present theme. It illustrates, first of all, the systematic application

of pure scientific principles to a desired and well-defined end which, though primarily commercial, is not without social implications. Second, the research threw light on the physico-chemical basis of anaesthesia, which time, if no other reason, forbids me to enlarge upon. Third, success was crucially dependent on the existence of a team of specialists in several scientific disciplines and also in the somewhat recondite technology of fluorine; notwithstanding these favourable circumstances and adequate financial backing, the research and development took seven years. Fourth, the distinction between the research and the development function of business is brought out: the original synthetic and early testing work was research, while the development consisted of devising suitable methods of manufacturing and launching the chosen compound, and of studying its clinical effect.

I now turn to the functions of design and construction, which are essentially the field of the technologist or engineer. His grasp of scientific principles, combined with the creative skill of a Telford or a Fowler, fits him above all to build a bridge between science and business. It has been said: 'The establishment of the class of engineers . . . is the more important because this class will, without doubt, constitute the direct and necessary instrument of coalition between men of science and industrialists, by which alone the new social order can commence.' These words of Auguste Comte remain as fresh and as true today as when they were written over a hundred and thirty years ago.

The production or operation function includes the application of technical methods often according to a routine, though it also demands much planning and organisation. Although such work is sometimes considered humdrum and repetitive compared with research and development, it is often essential for it to be done by highly qualified technical men. It is just as great a mistake to believe that unqualified men who come up what used to be called 'the hard way', are generally adequate, let alone best for production work, as to hold that good technical people are wasted in such jobs as plant management. On the contrary, they must be encouraged to regard it as an interesting and necessary part of their training for higher management.

If there are to be no weaknesses in any one part of the organisation, technical staff may have to be moved from one department to another, and this places a vital responsibility on those exercising the personnel function in selection, training and promotion; at the same time it is their duty to ensure that employees can find satisfaction in their work. This function also can be conveniently performed by a scientist, whose knowledge of what the technical departments require, helps him to place staff to the best advantage both of the business and themselves.

On the commercial side also, scientists may be needed in sales, sales service or purchasing, to achieve the best results for a business making technically complex products. I cannot agree that employment of technical men in commercial departments is necessarily wasteful: it gives flexibility within the business, and can often ensure smoother co-operation when a works manager or research chemist feels that his own, the technical point of view, is understood by a man in one of the commercial departments; moreover, as training for higher administrative posts, commercial experience is invaluable to the scientist.

When he reaches such a post, the scientist, recognising the value of science in business administration, will be well placed to exploit it. I am not thinking here only of new equipment in the office, such as electronic computers or apparatus for sensing letters and figures; science has more to offer than mere tools. It offers the whole philosophy of the scientific approach to the problems of management. I need only remind you of some of the steps involved: definition of the problem, observation and collation of the relevant facts, formulation of a hypothesis or principle, its testing and its general application.

This approach has been of untold value in the examination of human activity and its measurement, which is usually described as work study. It has also been very

successfully employed in the problems of supply and distribution, in adjustments of stock levels and in planning of delivery programmes. The full application of such methods usually involves some specialist techniques, especially of a mathematical kind, but even the first qualitative steps of the scientific approach can often lead directly to improvements in business efficiency. All the same, it is worth noting that an otherwise successful application of such techniques can founder on the inability to handle human relationships with proper understanding.

Finance has played a conspicuous part in my address and, in my experience, no technological business ever attempts to progress without continually seeking financial advice. However, I have failed to see in this country an equal acknowledgment that financial institutions need to take the benefit of continuing technological advice, though I am glad that some progress is being made in this direction overseas. For example, I know of a bank especially interested in natural resources which employs full-time geologists. Again, the World Bank organisation in Washington employs technical men not only at Headquarters, but also to carry out specific missions in various countries. Finance and technology will, I hope, become more and more interdependent as our businesses adapt themselves to current needs.

There is scope then for the scientist to infiltrate throughout almost the whole business structure. That being so, it is pertinent to ask, if scientists are to play an important part in business administration, is the balance between science and business likely to be upset? The answer to this question is, I think, threefold.

First, as British industry is at present constituted, the balance is heavily tilted the other way. That is to say, there are still many enterprises which are not employing qualified technical men even in tasks where one would expect to find them, and conversely there are no industries that appear to be harmfully dominated by scientists. This is due partly to the shortage of scientists and technologists with which we are all familiar, but partly also to historical circumstances. In the nineteenth century, due to this country's wealth of practical engineering experience, theoretical knowledge was undervalued. At the same time, an educational bias towards the classics, based on an aristocratic tradition, sometimes led the second and third generation of a manufacturer's family to forsake the business altogether for politics or the bar, or if not, at least to be quite unfamiliar with the value of science. Again, as we have seen, in some concerns scientists are relegated to the position of advisers to management and are allowed little part in the active control of the business.

Second, in businesses where scientists rightly play a prominent part in administration, non-technical men who nevertheless have a good understanding of scientific principles, are with advantage sometimes put into semi-technical jobs, and this helps to maintain a balanced partnership.

Third, and most important, the balance that we seek must come in great measure from within the individual himself, a balance provided by the education which the scientist and the non-scientist ought to receive. This, besides being central to my theme, is very much the personal concern of many of us here.

(*b*) *Balance in education*

There are three senses in which we might consider balance in education. The first arises from two different conceptions of the purpose of education. On the one hand there is the formation of character, and the development of intellect and personality. At the other extreme is the concept which I suppose finds its most powerful expression in the educational philosophy of Soviet Russia, that the purpose of education is to provide useful members of society in which the desires of the individual must be subordinated to the needs of society.

I do not believe that these two concepts are wholly conflicting nor that the second runs completely counter to the British tradition of individual liberty. A balanced

educational system should have room for both. The striking successes achieved by the Russians in science and technology are directly attributable to a system of education deliberately designed to produce large numbers of technically trained men and women with the ability to increase the economic power of the country through industrialisation. To this end they established a strong teaching profession to which recruitment is no problem because it enjoys the highest pay and status of all the professions. Though undoubtedly state direction still has a place in the Russian educational system, its purpose is being achieved to an increasing extent by incentives.

We are proud that our Universities, despite their growing financial dependence on the state, still retain their academic independence intact. Nevertheless, they have quite properly reacted to the stimulus of the country's needs which have led the older Universities, for so long the home of the humanities, themselves to enlarge their science faculties. The new St. Catherine's College at Oxford and Churchill College at Cambridge will give particular prominence to the study of science. The introduction of transfer scholarships to enable promising arts men to turn to science, has been readily accepted. More sixth-formers are reading science in Public and Grammar Schools. Some long-term financial incentives may operate here, but more important has been an intensive propaganda effort which I hope will be sustained by this Association's programme for organising lectures up and down the country on the lines of those so successfully pioneered by the Royal Institution in London.

So, in framing our formal education, we cannot entirely ignore the requirements of the society on whose prosperity depend the very conditions in which our educational system can progress; but our society will not be tolerable to live in, if its members are not permitted to develop their personalities and talents.

This brings me to the second sense in which I refer to balance in education. I am thinking here of the need for those who study the humanities to have some understanding of the workings of science, and for our scientists to widen their horizons by relating their science to its human context both historically and in terms of the present day. I have in the past on several occasions declared my belief that the gulf which at present separates our scientists and non-scientists, is much more easily bridged than is often imagined. For example, the qualities of mind which are fostered by studying chemistry and history are very much alike. Both chemists and historians must cultivate an inquisitive and acquisitive mind. Both need to exercise judgment, leavened with a good measure of scepticism in the exemplary tradition of Robert Boyle. Both must be imaginative, the chemist in advancing hypotheses to predict the future, the historian in doing so to explain the past.

The merits of a balanced curriculum are recognised in Soviet Russia, where children between 11 and 17 devote something like 46 per cent of their working hours to science subjects, with the emphasis predominantly on mathematics, and another 46 per cent to what we call arts subjects, with the emphasis here on Russian language and literature. Clearly a correct balance can never be achieved with premature specialisation, which the current demands for entry into many Universities impose on our most talented students.

There is a third and final sense in which I refer to balance in education. I have said that we demand, from our scientists and non-scientists alike, intellectual qualities, such as curiosity, judgment and imagination. Equally we want our young people, whatever their branch of study, to possess high qualities of character. I need not enlarge on this theme except to say that the society which a balanced partnership of science and business can help to build, will as always call for courage, integrity and tenacity. These are some of the ideals that we might have in our minds, as we each address ourselves to the problems of education for an age of science.

Conclusion

My theme, ladies and gentlemen, has been broad and general. It does not call for revolutionary action, it calls rather for sustained progress, and, of the two, that is sometimes the harder to inspire.

In education Glasgow has an enviable reputation. I have already referred to some of the famous names connected with the University. Alongside its fine academic tradition, we can rejoice in over 150 years' growth of higher technical education. The Royal College of Science and Technology, as its name has been since 1956, can trace its origins back both to Birkbeck's Mechanics' Institution and also to Anderson's College, founded in 1796, where Birkbeck himself taught.

Where such institutions exist to teach and practise science, and where there is still the business acumen and mercantile tradition of the Glasgow 'Tobacco Lords', I hope my plea for extending the partnership between business and science will fall on fruitful ground. In Glasgow as long ago as 1799, Tennant exploited the discovery of chlorine to found a large chemical business. Again, in this city was discovered the cyanide process for separating gold from its ores, a process which has enabled gold to this day to be mined on a scale massive enough for it to be still a measure of world finance and commerce. Symington's *Charlotte Dundas* and Bell's *Comet* were fore-runners in the technology of power-driven ships such as have made the Clyde shipbuilder's skill famous throughout the world. Kelvin's fundamental work on electricity and his maritime interests led him to found a business based soundly on practical physics.

I have tried to show why science and business must be partners: the advancement of science can only be achieved with the resources which business can provide and, if business is to provide them, then it must bring science to its aid. I have also suggested that a truly balanced partnership will manifest itself in the growing use of science and its methods in business; and that it can only come from a balanced education for scientists and business men alike.

I believe this balance to be important for our well-being, and as we inaugurate this 120th Meeting of the British Association, we can take with sincerity and all humility for our own work, the objective for his college expressed in Anderson's will:

'For the good of Mankind and the Improvement of Science.'

SOME ASPECTS OF THERMONUCLEAR RESEARCH[1]

ADDRESS BY DR. T. E. ALLIBONE, F.R.S.

PRESIDENT OF SECTION A

AT OUR DUBLIN MEETING last year Mr. Lawson and Sir George Thomson reviewed the way by which reactions between hydrogen nuclei might be caused so that useful power could be generated. From data on the reaction cross section for deuterium ions it is obvious that they must have energies of tens of kilovolts in order to have a reasonably high fusion probability and the ions must be maintained as a hot gas at temperatures of over 100 million degrees for quite long periods of time. Mr. Lawson drew our attention to the fact that such a hot gas does not lose energy by radiation according to the Stefan-Boltzmann law which applies to black bodies, and which depends on the fourth power of the temperature, but instead, and indeed fortunately for our present purpose, loses energy only according to a square root function of temperature. He spoke of a critical temperature for a volume of this completely ionised gas as that at which the rate of release of fusion energy just exceeds the radiation loss; for the deuterium D–D reaction, this T_c is of the order of 10^8 degrees and for the D–T reaction is three times lower. Even these temperatures are based on retaining the products of the reaction until their energy has been absorbed in the gas; if, however, some products escape, as in practice the uncharged neutrons must do, the critical temperature will be higher. If all the reaction products escape, the inefficiency inherent in converting their energy into electricity to maintain the system, sets a lower limit on the product of the density of the gas and the time for which it must be kept hot. In this case short transient methods are ruled out because the densities needed are unpracticable.

Sir George Thomson reflected upon the ways in which a hot gas might be held together without destroying the vessel containing it and quickly dismissed several nice ideas which cannot be made to work, ultimately coming down in favour of the pinch effect method in which the electromagnetic forces associated with a high current discharge pull the gas away from the walls and compress it into a narrow cylindrical channel. Such pinched discharges have instabilities and he suggested ways by which these might be controlled.

Since we met in Dublin there has been another year of intense activity in thermonuclear research and some measure of agreement between American and British Governments as to publication. The year has also seen the achievement of some discharges in deuterium carrying high currents for relatively long periods of time, held in reasonably stable positions, and in a number of these discharges neutrons have been produced copiously. To an experimental physicist in this field the problems which beset him are numerous; some he may think he solves; others he knows he has to pass over unresolved for the moment. The theoretician too is faced with immense complexities and rarely gets a clean-cut answer from experiments, probably because these have been pressed on with some superficiality, but progress there has been and it is my object to review briefly what has been achieved.

In this country all our efforts have been concentrated on the method of approach involving the pinch effect in a high current discharge, and in other countries too this method has received the most attention. Other methods are being studied; thus for example the American Atomic Energy Commission recently announced three main projects involving other methods.(1) In the work at Livermore and

[1] Address delivered on September 1, 1958, at the Glasgow Meeting of the British Association.

Oakridge laboratories the hot gas is contained essentially in a cylindrical vessel with a very large magnetic field running parallel to the axis (Fig. 1). This field makes the radial diffusion and heat conduction in the gas very small. Losses in the longitudinal directions are prevented by making the magnetic fields at the ends of the cylinder still larger and the resulting constrictions of the lines of forces at these points act like magnetic mirrors and reflect charged particles which approach the ends. Such devices are known as 'Mirror Machines' or 'Pyrotrons' and their field configuration will contain a hot gas effectively provided the number of particle collisions is small. Fortunately this is the case at temperatures of 10^8 degrees but it is not true at low temperatures. In this method therefore the experimenters aim at generating initially a very high temperature but low density gas, and then raising the density. This is in contrast to the pinch effect approach which starts with the required density and raises the temperature.

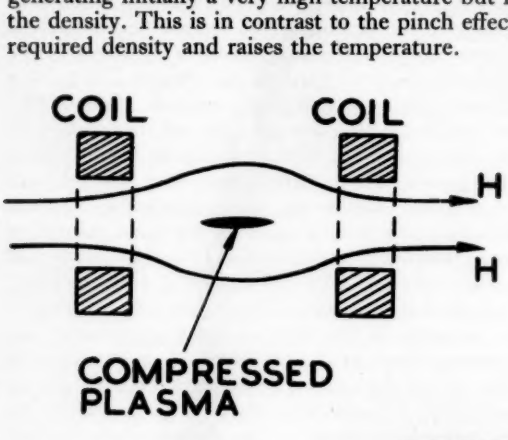


Fig. 1.—Magnetic configuration for a mirror machine.

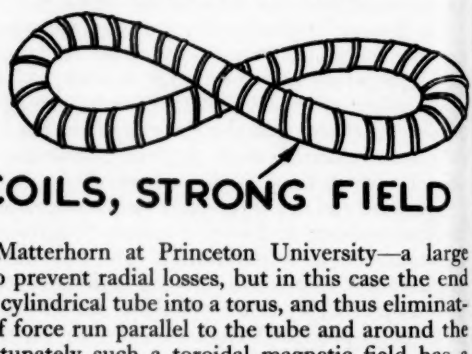


Fig. 2.—Shape of the Stellerator discharge tube.

In the third project—Project Matterhorn at Princeton University—a large magnetic axial field is again used to prevent radial losses, but in this case the end losses are overcome by bending the cylindrical tube into a torus, and thus eliminating the ends. The magnetic lines of force run parallel to the tube and around the circumference of the torus. Unfortunately such a toroidal magnetic field has a larger magnitude near the inside wall of the torus than near the outside wall. The variation in magnitude would cause a transverse diffusion of the charged particles, and this effect has been overcome in the 'Stellerator' by twisting the torus into a figure-of-eight shape (Fig. 2).

In addition to the different magnetic field configurations which are being used to achieve containment of the hot gas, there is also a choice in the way in which to heat the gas. The various methods are ohmic heating due to electric currents in the gas, adiabatic compression, heating caused by alternate irreversible compressions and expansions, shock waves, and the injection of energetic charged or uncharged particles. Ohmic heating is generally associated with the pinch effect approach, and

at first sight it looks the most suitable method of dissipating very large powers in the gas. The snag is that the electrical conductivity varies as $T^{3/2}$ and at 10^8 degrees it is twenty times better than copper, and hence very large currents are needed to generate a required power. It is because of this that other methods are being actively studied. In the case of the Mirror Machines the main heating method being studied is the injection of fast particles since here energetic gas particles are required from the start.

Unfortunately no details have been published yet of the methods of approach other than the pinch effect, though we can expect them at the Geneva Conference which is starting today. Because of this I will concentrate on giving you details of the pinch effect method. This is justified since it is the main method being studied and because it is the method which has yielded such interesting results in the past year.

We owe to Sir George Thomson the original concept of using the pinch effect in a gas discharge in a toroidal vessel as a way to contain the hot plasma and keep it from the walls of the vessel. By making the discharge curl up on itself in such a torus, end effects at electrodes are avoided. During the war Steenbeck in Germany had produced such a toroidal discharge carrying a few thousand amperes, but his goal was to produce high-speed runaway electrons as in a betatron, not to achieve high temperatures as such. Thomson set Alan Ware to work on this problem in 1947 and in a short time Ware had generated currents of 30,000 amperes in a 16-in. diameter glass torus having a bore of $1\frac{1}{4}$ in. and was the first to show a marked pinch effect in a gas discharge.[2] He was using a 2,000 joules bank of condensers, and examined the discharge with rotating mirror camera, thus (Plate I) showing it to pinch to narrow diameter in 2 or 3 microseconds and then expand again, from which he concluded that the inertia of the rapid contraction was causing the discharge to constrict beyond the equilibrium position and hence 'bounce' out again. Further contractions and expansions followed. When the energy fed into the discharge was increased spectroscopic observation showed that the walls were evaporating as the discharge struck them, and both glass and quartz tori evaporated.

In Oxford Dr. Thonemann was conducting similar experiments on gas discharges independently and in 1951 his work was transferred to A.E.R.E. Harwell, and Ware's to A.E.I. Aldermaston. Since that date the two teams have been built up working in close collaboration, the work in Aldermaston having been supported by a Government contract, and the work generally planned on a joint basis. As far as we know there had been no similar work in America or Russia up to this date.

In 1951 the first metal torus was made in Aldermaston in aluminium to try to avoid the troubles due to fracture of glass and to conduct away the heat pulse to avoid evaporation. Two gaps were provided in this first torus to prevent the metal tube from acting as a short-circuited turn, and they were bridged by insulating rings shielded in some measure from direct radiation from the discharge channel. Discharges of many thousands of amperes in hydrogen were produced but arc discharges bridged the gaps and many attempts were made to avoid this. Tori having very many porcelain spacing rings were tried, and one torus, with tube bore 30 cm. and torus diameter 100 cm., tested in 1955 during the Geneva Conference, carried 100,000 amperes in deuterium gas for two hundred microseconds (Plate II). The discharge pinched to narrow diameter and remained so but was subject to violent lateral movements. Counting apparatus was set up but no neutrons were detected.

In 1954 Kruskal and Schwarzschild derived a theory of the instabilities which must arise in a gas discharge pinched by its self-magnetic field. If any kink develops due to any geometrical or electrical irregularity, the surrounding magnetic field exerts a greater pressure on the inside of the kink than on the outside, the pressure varying inversely as the radius of the kink curvature and the kink grows

in amplitude. Carruthers and Davenport (1957)[3] had observed these instabilities in A.E.R.E. a year before this theory was published: they had been using straight tubes and also a torus 30 cm. diameter, 10 cm. bore, pulsed for 500 μs, with currents in argon and krypton, and observed the discharge with an image-converter camera at different moments during the pulse. They showed that after the current channel has pinched it becomes unstable and assumes a rapidly fluctuating helical configuration having a helical pitch several times the dimension of the bore of the tube. Hemmings, Miles and Ware (Venice, 1957)[4] showed that in argon the instability grows with a velocity of 5×10^5 cm./sec. (Plate III). In hydrogen they grow at a faster rate, and at high currents velocities up to 10^7 cm./sec., have been observed.

Unknown to us in England at that time, Artsimovich with collaborators in Russia had been experimenting from 1951 with straight discharge tubes, and so

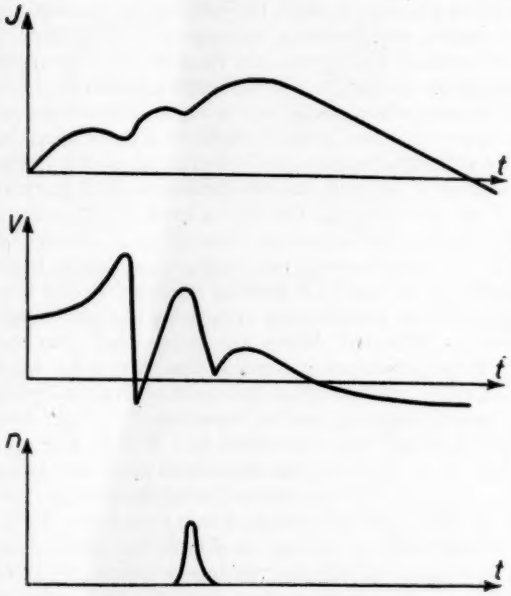


Fig. 3.—Diagrammatic representation of the time variation of current and voltage in linear discharge, showing also the burst of neutrons. (Kurchatov, *J. Nuclear Energy*, 2, 1957.)

had a number of American scientists. Straight tubes have the experimental advantage that with them it is much easier to obtain high currents and high rates of change of current than with toroidal tubes. They have the disadvantage that there is no heat insulation from the end electrodes and after a short period metal vapour must enter and quench the discharge. Because of this such tubes are never likely to be used in a final thermonuclear reactor. The Russian investigators used transparent discharge tubes, porcelain tubes, and metal tubes up to 60 cm. diameter, and up to 2 m. long, and passed currents up to 2 million amperes in various gases and with rates of rise of current up to 10^{12} A./sec. and used a bank of condensers conserving up to 250,000 joules. These currents and rates of rise of current were much greater than we had used. They had observed (by 1956, April, the date of Kurchatov's lecture in Harwell)[5] (Plate IV) rapid pinching of the discharge followed by expansion, then further contractions and expansions, and finally blowing up of the discharge. The associated variation of inductance caused corresponding fluctuations in the discharge voltage (Fig. 3). Quite early in these experiments, in 1952 according to Kurchatov, neutrons were found in deuterium

with gas pressure of 0·1 mm. and at currents over 400 kA., the neutron yield increasing rapidly with peak current. At first these were naturally thought to be due to D–D reactions in the gas in equilibrium at high temperature as expected with such large currents, but doubt was cast on this when later the investigators found they could detect neutrons at 150 kA., and Artsimovich later showed that the yield, rising ten times between 220 and 400 kA. ceased to rise with higher currents. The neutrons appeared as bursts of about 10^7–10^8 at the second and third kinks of the voltage oscillograms, and were eliminated if 0·01 per cent of gas impurity was added to the deuterium. Although their source was somewhere along the length of the gas discharge and not concentrated near to the cathode, a study of the energy of the neutrons showed that the reactions were caused by deuterons with energies of 50–100 keV directed parallel to the tube axis. This proved the reactions were not thermonuclear in origin.

In very similar work at the Radiation Laboratory of the University of California which started at about the same time but was not published until the Venice Conference of 1957,[6] much the same results were obtained. The research workers

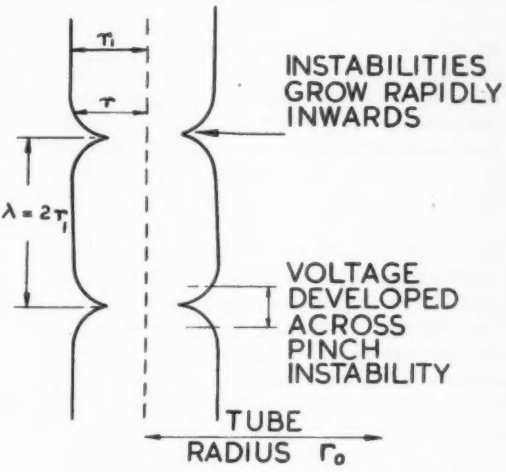


Fig. 4.—Diagrammatic representation of the sausage type instability. (From drawing by Anderson *et al.*, Venice Conference, 1957.)

there also went through the stages of initial exaltation, growing doubts, and finally experimental proof that the reactions were caused by directional acceleration of some of the deuterons. Various theories have been put forward to explain how some of the particles are accelerated to energies greatly in excess of the applied voltages. The most convincing is that due to Colgate. An extension of Kruskal's theory shows that there is one mode of instability in which the discharge over-constricts at some places along its length and expands at the points in between (Fig. 4). This instability is known as the 'sausage' instability, and the rapid change of inductance at the points of constriction can cause large local voltages. A few of the particles can traverse these regions and gain high energies. Some photographic evidence that this form of instability exists in practice has been obtained at the University of California.

By 1954 it was clear to us in this country that there were two main problems which had to be solved before the pinch effect approach could be expected to yield high temperatures; one was the discharge instability and the other the vaporisation of the tube walls. Of these the former was recognised to be the more fundamental, since no material could be expected to withstand the bombardment caused by the close approach of an unstable and hot discharge. On the other hand the development

of suitable materials to withstand the radiation and limited bombardment from a stabilised and well contained discharge could be expected. In the subsequent years many experiments were carried out with extra electromagnetic fields applied to the discharge in order to control the instability. The most obvious field to try is a steady toroidal magnetic field running parallel to the discharge since many earlier workers with low current arcs had found that axial fields anchored the discharge in one position. The experiments however presented a picture far from simple (Hemmings *et al.*, *loc. cit.*). In many cases the magnetic field destroyed the pinch effect and in other experiments at lower fields there was evidence of increased instability.

With further experiments our understanding of the behaviour of such discharges in the presence of magnetic fields improved, particularly as a result of the contributions of Bickerton and Tayler at A.E.R.E. and Liley at A.E.I. It was realised that an axial magnetic inside the discharge would give the discharge stiffness and prevent the growth of sausage instability and short wavelength wriggle instabilities. On the other hand an appreciable axial magnetic field on the outside of the discharge is detrimental in that it encourages the growth of certain modes of the wriggle instability. A field configuration was therefore required in which there was an axial magnetic inside the discharge but little or none outside. Lastly, in order to achieve complete stability it was necessary to combine this configuration with a metal tube whose eddy currents would suppress the long wavelength wriggle instabilities. The exact requirements for stability came as a result of careful mathematical analysis by Tayler in 1956 (see Tayler, 1957)[7]. Rather surprisingly the same results were obtained independently at almost the same time by Rosenbluth in America (1957)[8] and by Shrafranov in Russia (1956)[9].

Fortunately, Nature has made it fairly easy to obtain the magnetic field configuration required for stability. This is because, if a discharge is created in a metal tube with an axial magnetic field, the radial contraction of the discharge across the field causes currents to circulate in the gas which enhance the magnetic field within the discharge. This increase of magnetic field causes further currents to circulate in the opposite direction in the metal tube and the magnetic field of these currents cancels the applied magnetic field outside the discharge. This effect is known as field 'trapping' since the contracting gas appears to carry the magnetic lines of force in with it.

The first published experimental work describing the observation of magnetic field trapping was that of Bezbatchenko *et al.* described at the 1956 Stockholm Conference.[10] Fields up to 12,000 gauss were applied to discharges of the type already referred to by Artsimovich. Four main effects may be mentioned: (*a*) the degree of contraction (Plate V) of the plasma diminishes as the applied field is increased in value; (*b*) the rate of contraction diminishes, and oscillograms show that the rapid rise of potential across the tube due to the rapid increase of inductance no longer occurs when fields of 8,000 gauss are imposed; (*c*) after the initial radial oscillations of the plasma, without the field the discharge rapidly blows up and reaches the walls, but with the field the expansion is retarded; and (*d*) measurement of the flux with magnetic probes shows that the conductivity of the plasma in the first stage of the pinch, say for 3 μs, is insufficient to trap all the longitudinal lines of force, but rises rapidly—by two orders of magnitude—in the following 3 μs during which no further escape of the field occurs. All the Russian work reported concerns discharges lasting for only a short time; after 6 μs the peak current has been passed and records are not available. No reference is made to the effect on the neutron yield.

The American investigators (Anderson *et al.*, *loc. cit.*) applied up to 100 gauss longitudinally and report only the effect on the neutron yield; the yield falls off inversely with field and is zero at 100 gauss; calculation showed that the internal

Plate I.—Streak photograph of 15,000 ampere discharge in low-pressure krypton, showing contracting and expanding discharge. The tube diameter is 1¼ in.
(Ware, *Engineering*. November, 1957.)

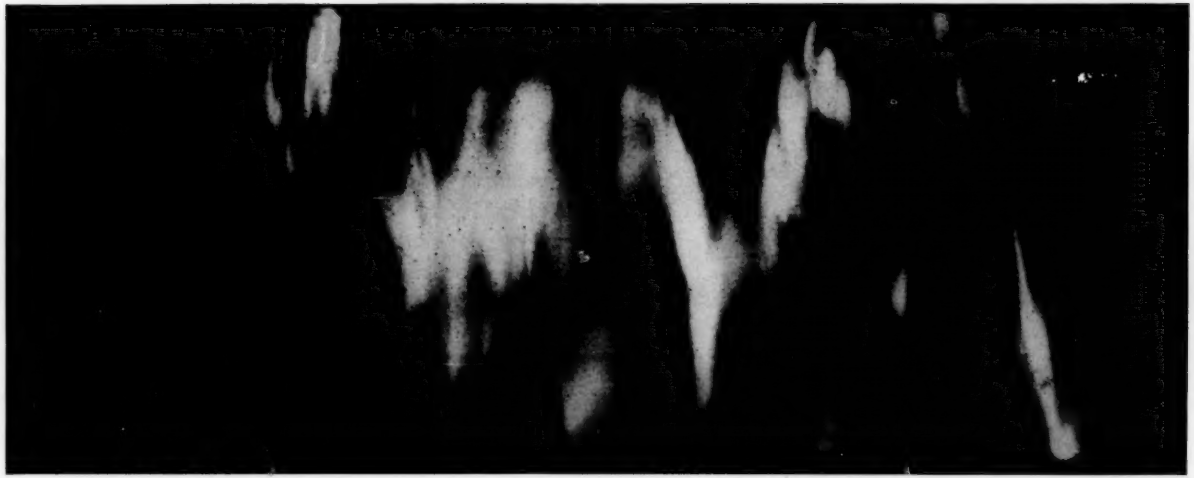


Plate II.—Streak photograph of an 80,000 ampere discharge in heavy hydrogen, showing violent wriggling of discharge. The tube diameter is 12 in.
(Ware, *Engineering*. November, 1957.)

Plate IV.
Hemisph

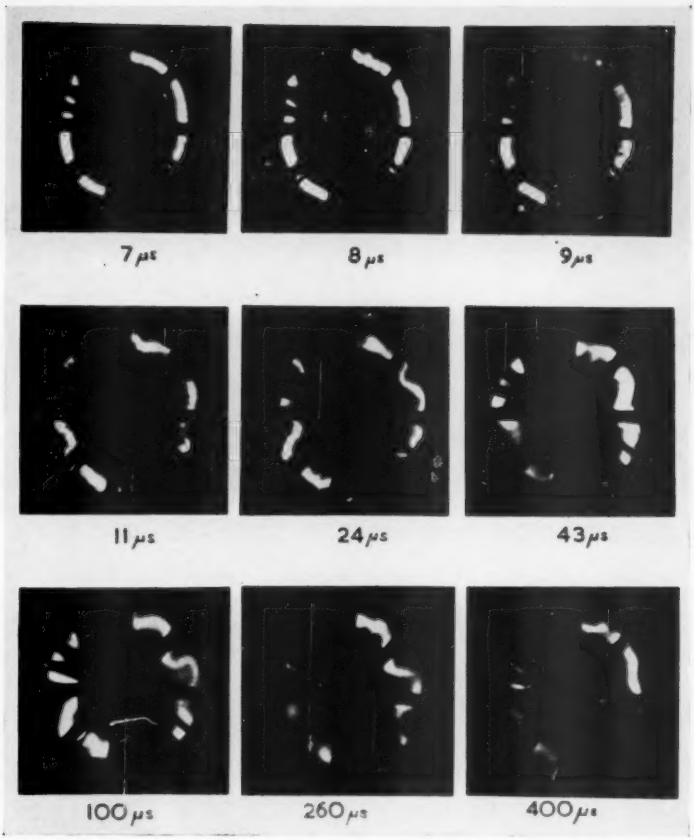


Plate III.—Image converter photographs of an argon discharge taken at different times during the pulse. (One microsecond exposure.)
(Hemmings, Miles, Ware, Venice Conference, 1957.)

Plate IV.—Photographic sweep of a discharge in deuterium at a pressure of $P_0 = 10$ mm. Hg. Hemispherical electrodes were held 45 mm. apart in a 180-mm. diameter chamber. $I_{max} = 1{\cdot}2 \times 10^6$ amps, $T/4 = 9{\cdot}5$ μsec. Scale 18 mm.=1 μsec.
(Kurchatov, J., *Nuclear Energy*, **2.** 1957.)

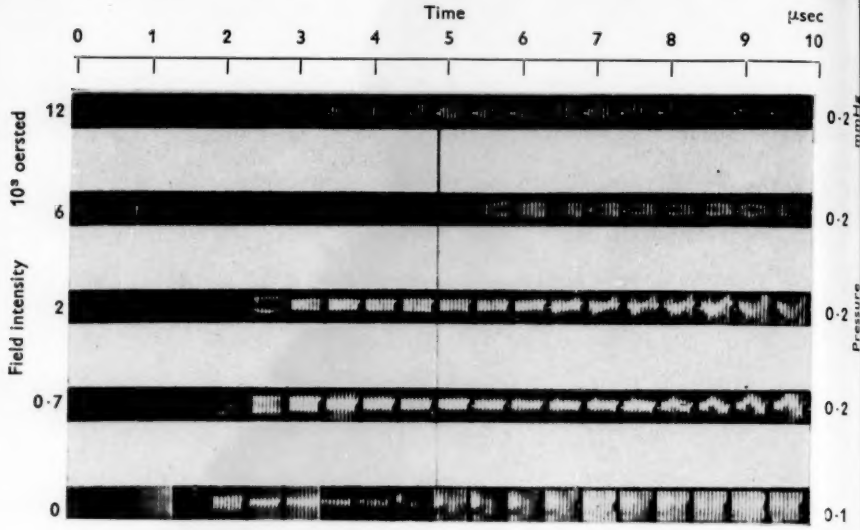


Plate V.—High-speed photographs of linear discharge with applied magnetic field. The dark lines across the discharge are the shadow of the coil which produces the longtitudinal magnetic field.

(Bezbatchenko *et al.*, *J. Nuclear Energy*, **2.** 1957.)

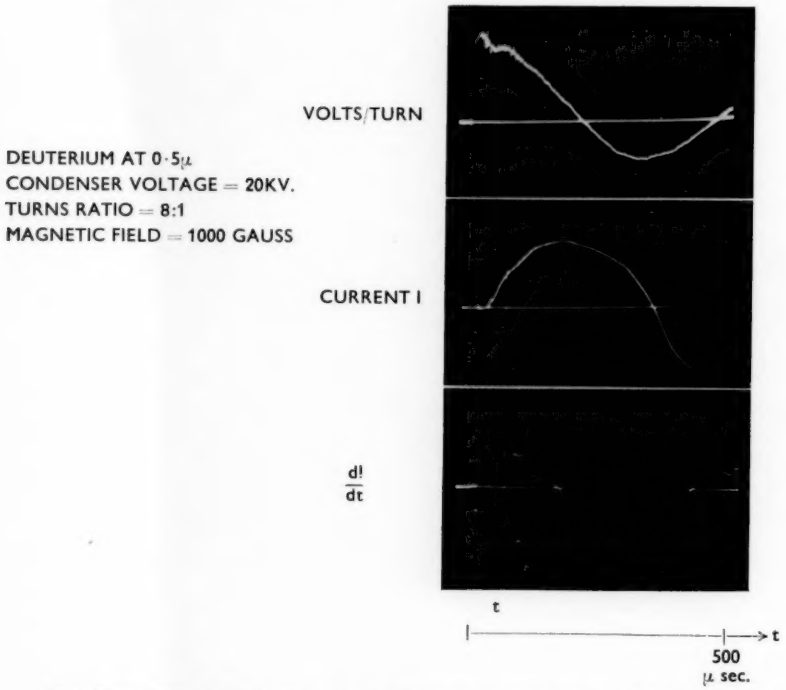


Plate VI.—Example oscillograms of faults per turn, current and rate of change of current for SCEPTRE III.

(Allen, Allibone *et al.*, *Nature, Lond.* January, 1958.)

pressure of the field prevented the discharge from constricting to the radius necessary for the longitudinal accelerating voltages to reach high value. Most recent results in America (Burkhardt *et al.*, 1958,[11] Hagerman *et al.*, 1958[12]) confirm that in straight tubes carrying currents of 10^6 A. for a few μsecs., fields of 1,000 gauss prevented pinch instability but also greatly reduced neutron yields: small amounts of impurities also reduced neutron yield—it will be seen later how this stands in marked contrast with results on long-duration discharges.

In England, a discharge with the trapped magnetic field configuration was first observed by Bickerton in 1956 (Venice 1957).[13] With a metal torus (diameters 30 cm. and 10 cm.) large currents were studied in hydrogen lasting for hundreds of microseconds in the presence of axial fields of several thousand gauss. Magnetic probe measurements indicated that the discharge was stable. After 170 μs, for

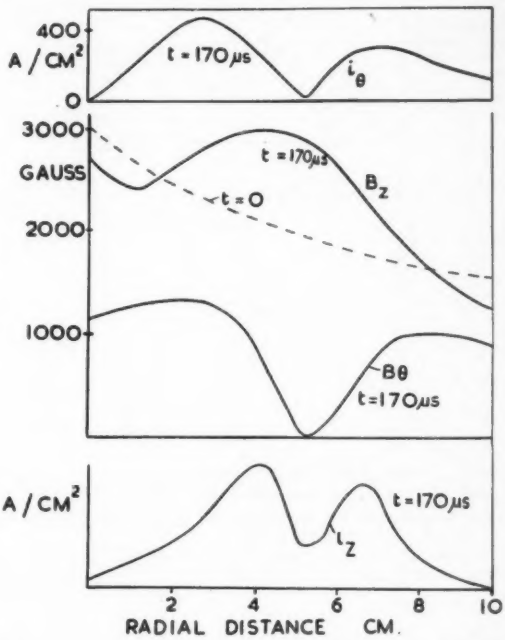


Fig. 5.—Plots of current density and magnetic field against radius for a toroidal discharge in a metal torus. Taken 170 microseconds after the start of the pulse.
From drawing by Bickerton, Venice Conference, 1957.)

example (Fig. 5), the axial field is seen to be greater than the applied field in the centre of the discharge and less at the walls of the torus, and from the measured value of the self-magnetic field of the discharge, the current distribution over the cross section of the discharge can be derived; this shows the current to be carried on a hollow cylinder of about one third of the diameter of the torus. A plot of the contracting force $i_z B_\theta$ and the outward force $i_\theta B_z$ (Fig. 6) shows that they are roughly equal and opposite at all radii. Unfortunately the stabilisation theory of Tayler (*loc. cit.*) shows that the discharge of 27 kA. should be pinched but not stable; however, theory makes several simplifying assumptions such as infinite electrical conductivity, which are not true in practice. In this particular case the enhanced magnetic field in the discharge may have been due to a paramagnetic effect in the gas rather than to field trapping.

In August 1957 in the apparatus known as ZETA at A.E.R.E. neutrons were detected for the first time from a toroidal discharge (Thonemann *et al.*, 1958),[14] and shortly afterwards similar results were obtained with SCEPTRE at A.E.I. (Allen

et al., 1958),[15] and PERHAPSATRON at the University of California, Los Alamos (Housaker *et al.*, 1958).[16] The sizes of the toroidal tubes varied widely in these experiments, thus ZETA is an aluminium tube with bore 100 cm. and torus diameter 300 cm., SCEPTRE is an aluminium tube with diameters 30 cm. and 115 cm., and PERHAPSATRON is a metal sheathed quartz tube with diameters 5 cm. and 32 cm. In all three cases the results showed that stabilised discharges had been obtained with trapped magnetic field configurations. In ZETA and PERHAPSATRON magnetic probe measurements showed the enhancement of the axial field within the discharge and reduction near the walls, and the measurements were free from the major fluctuations associated with instability. With SCEPTRE, measurements of

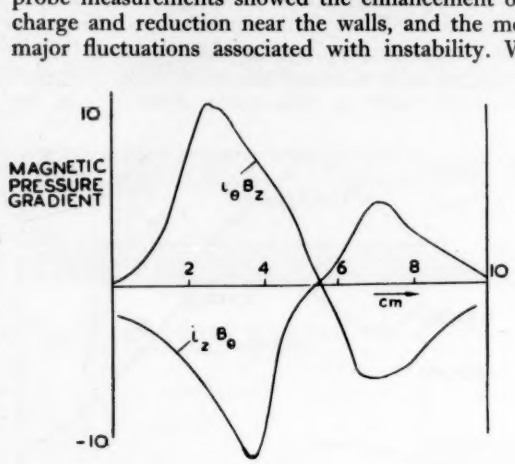


Fig. 6.—Plots of the products $i_\theta B_z$ $i_z B_\theta$ against radius for a metal torus.
(From drawing by Bickerton, Venice Conference, 1957.)

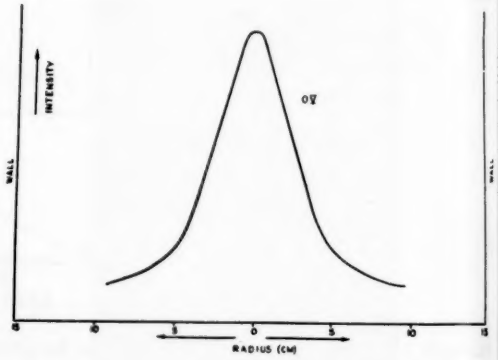


Fig. 7.—Variations of the light from quadruply ionised oxygen across the tube at different times during the discharge with SCEPTRE III.

inductance from the rate of change of current indicated the discharge leaving the walls and constricting when the self-magnetic field equals the applied magnetic field, as expected when trapping of the field occurs (Plate VI). In addition a photoelectric time- and spatially-resolved study of the light from highly ionised impurity atoms showed the light came from a narrow core which always remained in the centre of the tube (Fig. 7). With ZETA and SCEPTRE the discharge was held stable for long periods, namely 2 and 0·5 milliseconds respectively.

The currents studied in these experiments were in the range of 5×10^4 to 2×10^5 A., and in all three cases neutrons were observed at higher currents. The nature of the neutron emission was very different from the early work with linear discharges. Thus neutrons were emitted throughout most of the current pulse and not in short bursts. An axial magnetic field was necessary for neutrons and in each case maximum yields were obtained at approximately that magnetic field which

would limit the discharge constriction to one third of the tube diameter. With 25 per cent nitrogen impurity in ZETA some neutrons were still detected and with SCEPTRE 10 per cent argon produced only a small reduction in yield.

In general the results with the three experiments follow simple scaling laws which confirm the theory of the stabilised discharge. Thus the time constants are proportional to the square of the tube radius (b^2), the pressures proportional to $1/b^2$. and magnetic fields proportional to $1/b$. The number of reactions per unit volume is proportional to $1/b^2$, whereas the number of reactions per unit length of tube is independent of b. Since in addition the energy input per unit length is independent of b the three experiments are equally efficient in producing neutrons.

With ZETA and SCEPTRE measurements were made to determine the ion temperature in the gas. This was done by observing the Doppler broadening of the spectral lines of impurity ions in the discharge, e.g. quadruply ionised oxygen (Fig. 8). The shape of these lines was found to be consistent with the atoms having a Maxwellian

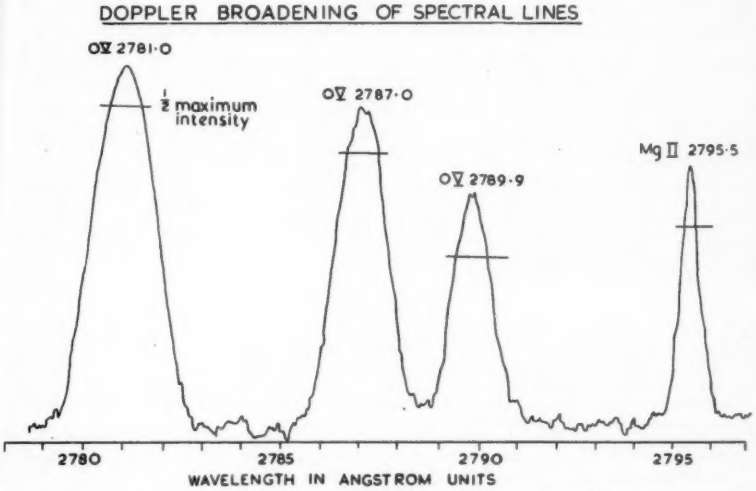


Fig. 8.—Microphotometer trace of part of a discharge spectrogram, showing broadened OV lines and also a narrower Mg II line from a cooler part of the discharge. The intensity scale is not linear: the half intensity point is marked on each line.

distribution of velocities and a temperature of several million degrees. It was not proof of the temperature, however, since a random instability or wave motion in the gas could have mass velocities which, averaged over time, give a Maxwellian distribution. Nevertheless these ion temperatures were found to agree reasonably well with those which would be needed to produce the observed neutrons by thermonuclear reactions. This fact together with the marked differences in the observed neutrons compared with the linear discharge made the scientists involved fairly confident that they had observed true thermonuclear reactions. Further confirmation was given by the fact that a theory of the ohmic and adiabatic heating by Liley at A.E.I. predicted temperatures of the order observed (see Allen *et al.*, *loc. cit.*). Such was the state when the announcements were made in January of this year.

In the subsequent months further measurements began throwing doubts on these conclusions. In many cases the Doppler broadening and neutron 'temperatures' were found to differ quite widely and what evidence there was of the electron temperature suggested that it was below a million degrees. The apparent ion

temperatures indicated a gas kinetic pressure comparable with the magnetic pressure, and this was inconsistent with the inductance measurements. Following this the scientists at A.E.R.E. were successful in performing the difficult experiment of measuring the neutron energies (Rose *et al.*, 1958).[17] This showed conclusively that the neutrons were not thermonuclear! The reactions were being caused by a few of the deuterons being accelerated parallel to the discharge current to energies of 17 keV. Once again, therefore, Nature appears to have fooled the scientists, this time in a rather more subtle way.

This discovery has been a great disappointment to the scientists concerned. I must point out, however, that the evidence for a stable discharge maintained for a long period still stands, and this in itself is a major advance; just over a year ago the instability problem was the main problem holding up progress. The big questions which now have to be answered are—what are the true ion and electron temperatures in ZETA and SCEPTRE and how is the heat generated in gas being lost to the walls?

There are several features of the results which are not understood; for example high speed photography shows bright bars of light travelling along the tube. These suggest some form of wave motion and they may well present a mechanism for heat loss to the walls. Again arc breakdowns still occur at the gaps in the metal torus in the case of ZETA and SCEPTRE, although it is much less severe than in the presence of an unstable discharge. This arcing is thought to be the cause of the impurities observed in the discharge, and the arc spots which travel along the inside surface of the tube may provide paths for heat conduction losses to the walls. Perhaps the next advance will come through improved materials for the tube walls. If the arc breakdowns can be avoided—and we are conducting research to achieve this—the major heat losses may be removed.

References

(1) American Physical Society Meeting, Washington, May 3, 1958.
(2) Cousins, S. W., & Ware, A. A. (1951): *Proc. Phys. Soc.* **64B,** 159.
(3) Carruthers, R., & Davenport, P. A. (1957): *Proc. Phys. Soc.* **70B,** 49.
(4) Hemmings, R. F., Miles, H. T., & Ware, A. A. (1957): *Proceedings of Third International Congress on Ionisation Phenomena in Gases*, p. 464.
(5) Kurchatov, I. V. (1956): *Atomnaya Energiya*, **1,** and five other papers in the same journal.
(6) Anderson, O. A., Baker, W. R., Colgate, S. A., Ise, J., & Pyle, R. V. (1957): *Proceedings of Third International Congress on Ionisation Phenomena in Gases*, p. 62.
(7) Tayler, R. J. (1957): *Proc. Phys. Soc.* **70B,** 1049.
(8) Rosenbluth, M. N. (1957): *Proceedings of Third International Congress on Ionisation Phenomena in Gases*, p. 903.
(9) Shrafanov, V. D. (1956): *Atomnaya Energiya*, **1,** 5, 38.
(10) Bezbatchenko, A. L., Golovin, I. N., Ivanov, D. P., Kirillov, V. D., & Yavlinsky, N. A. (1956): *Atomnaya Energiya*, **1,** 5, 26.
(11) Burkhardt, L. C., Lovberg, R. H., & Phillips, J. A. (1958): *Nature, Lond.*, **181,** 224.
(12) Hagerman, D. C., & Mather, J. W. (1958): *Nature, Lond.*, **181,** 226.
(13) Bickerton, R. J. (1957): *Proceedings of Third International Congress on Ionisation Phenomena in Gases*, p. 101.
(14) Thonemann, P. C. *et al.* (1958): *Nature, Lond.*, **181,** 217.
(15) Allen, N. L., *et al.* (1958): *Nature, Lond.*, **181,** 222.
(16) Housaker, J., *et al.* (1958): *Nature, Lond.*, **181,** 231.
(17) Rose, B., Taylor, A. E., & Wood, E. (1958): *Nature, Lond.*, **181,** 1630.

THE TRANSURANIC ELEMENTS[1]

ADDRESS BY PROFESSOR H. J. EMELÉUS, C.B.E., F.R.S.
PRESIDENT OF SECTION B

THE DISCOVERY of nuclear fission and of the transuranic elements is without question the most significant advance made throughout the history of science. We have yet to see the full benefits which will accrue to mankind, though the pattern is already beginning to emerge in current programmes for the production of nuclear power and in the ready availability of radioactive isotopes for therapeutic use. Most credit for the progress of the last fifty years goes to the physicist, but the chemist has played an important part and the future will involve increasingly close co-operation between physicists, chemists and metallurgists. It is not inappropriate, therefore, to consider today the bearing of this new field on progress in chemistry.

The actual isolation of the transuranic elements was the crowning achievement of a series of discoveries spread over about sixty years. For the first of these we have to go back to the close of the last century when X-rays were discovered in 1895 and when Becquerel, in seeking a connection between X-rays and the then familiar phosphorescence of uranium salts, discovered that if a sample of potassium uranyl sulphate was wrapped in black paper and placed on a photographic plate, the latter was blackened. Becquerel also observed that the penetrating rays from uranium could discharge an electroscope. Pierre and Madame Curie in 1898 concluded that the penetrating rays from uranium were a new atomic phenomenon and it was they who introduced the name radioactivity.

The subsequent history of the subject is so familiar that I need refer to it only very briefly. It was soon established that thorium was also radioactive. Natural uranium ores were found to be more radioactive than would be expected from their uranium content and it was the chemical investigation of this anomaly which led to the isolation of polonium and radium by the Curies. In the meantime, the nature of the rays from radioactive substances was elucidated and in 1903 Rutherford and Soddy introduced the disintegration hypothesis, which enabled much of the new knowledge to be put on a systematic basis.

The next phase was spread over some twenty years, during which the minute nucleus of the atom and its planetary electrons became established as physical realities. Radioactivity was finally confirmed as a nuclear property and it was the study of atomic nuclei which started the series of advances culminating in the discovery of fission and the transuranic elements. The first of these was the observation of the disintegration of the nuclei of some of the lighter elements when bombarded with the fast particles (α-particles) emitted in the decay of naturally radioactive elements. Similar nuclear disintegrations were next shown to be caused by particles (protons or helium ions) which had been accelerated artificially to velocities comparable with those of the particles from radioactive decay. The neutron was discovered in 1931 and, shortly afterwards, it was found that bombardment of the nuclei of lighter elements with α-particles could produce new nuclei which were radioactive. The same was found to be true for artificially accelerated particles and for neutrons of suitable energy.

The production of radioactive nuclei by neutron bombardment was of special significance for, since the neutron is uncharged, it is able to approach the nucleus without encountering the repulsive forces which operate for a proton or α-particle.

[1] Address delivered on August 28, 1958, at the Glasgow Meeting of the British Association.

Fermi, who first used neutrons in this way, was able to bombard heavy nuclei and obtain reactions such as that for gold shown below.

$$Au_{79}^{197} \xrightarrow{n,\gamma} Au_{79}^{198} \xrightarrow{\beta} Hg_{80}^{198}$$

The first reaction is the capture of a neutron and the emission of a γ-ray. The new nucleus is unstable and loses a β-particle to form a stable mercury isotope. The important feature about this process is that the product has an atomic number one greater than the element taken initially. Needless to say, the yield of the new product in such a case is very small and a very large proportion of the gold would normally be unchanged.

Fermi realised that if such a process were to operate for uranium (Atomic Number 92) the product would have an atomic number of 93 and be a new transuranic element. When the experiment was done with natural uranium a number of radioactive species were observed, each characterised by its mode and energy of decay. At first these new radioactive substances were thought to be transuranic elements, but it was established later that two distinct processes take place. The first is the fission of U^{235}, which yields radioactive isotopes of a number of the light elements. The second is neutron capture by U^{238}, the more abundant of the two chief uranium isotopes.

$$U_{92}^{238} \xrightarrow{n,\gamma} U_{92}^{239} \xrightarrow[23\ \text{min.}]{\beta} Np_{93}^{239} \xrightarrow[2{\cdot}3\ \text{days}]{\beta} Pu_{94}^{239} \xrightarrow{\alpha}$$

The first step is exactly analogous to the example cited above in which gold is bombarded. The product, U^{239}, undergoes two successive β-decays forming in turn neptunium and plutonium. These are the first of the series of ten transuranic elements listed below:

Z			Z		
93	Neptunium	Np	98	Californium	Cf
94	Plutonium	Pu	99	Einsteinium	Es
95	Americium	Am	100	Fermium	Fm
96	Curium	Cm	101	Mendelevium	Mv
97	Berkelium	Bk	102	Nobelium	No

There are, of course, various other ways of producing the isotopes of these elements, over a hundred of which are now known. These will not be dealt with in detail, though we may note the following as the chief types of reaction used:

(1) Bombardment with particles accelerated in the cyclotron or a similar high-voltage machine, e.g.

$$U_{92}^{238}\ (d, 2n)\ Np_{93}^{238}$$

$$Pu_{94}^{239}\ (\alpha, n)\ Cm_{96}^{242}$$

(2) Neutron bombardment, e.g.

$$Am^{241}\ (n, \gamma)\ Am^{242} \xrightarrow{\beta} Cm^{242}$$

(3) Multiple neutron capture.

(4) Bombardment with heavy ions.

It is not necessary to say much about the first two of these general methods. The first has been widely used and it is interesting that the product depends often

on the energy of the particles. The target itself may be an artificial element available in only small amounts. Thus, californium may be made from curium by the reaction Cm^{244} (α, $2n$) Cf^{246}, or einsteinium from californium by the reaction Cf^{249} (d, $3n$) Es^{248}. In all such cases a good deal of chemical manipulation is involved in separating the product from target and also from fission products, which are not infrequently formed simultaneously. For elements beyond curium the ion-exchange column, which is referred to later, is the main means of separation, but for an element such as plutonium which has several valency states, chemical reactions on a micro or even on a macro scale are often used.

Multiple neutron capture is possible only with high neutron fluxes, such as are available in some reactors. The figure below illustrates what happens to Pu^{239} under intense neutron bombardment. The horizontal steps are neutron captures and the vertical ones β-particle emission, with production of a unit increase in the nuclear charge. In a high-flux multiple neutron capture occurs until, at a

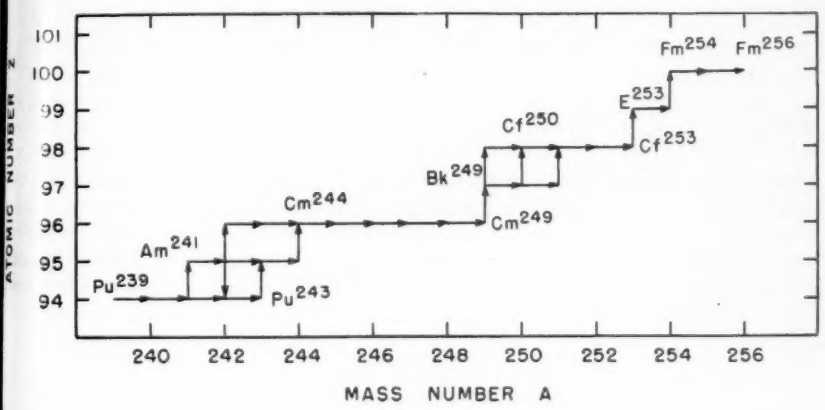


Fig. 1

(Reproduced by courtesy of Methuen & Co., Ltd., from *The Chemistry of the Actinide Elements* by Seaborg and Katz.)

certain point, a β-particle is emitted. In this case also ion exchange resins give a dependable means of separating and identifying the products. Sufficient is known about the stability of nuclei for the incidence of the steps to be predicted.

The use of heavy ions is a comparatively recent development. The cyclotron or a similar machine is used but the particles which are accelerated are nuclei such as Be^{9}, C^{13} or O^{16}. These lead to a large one-step increase in mass and reactions such as U_{92}^{238} (N^{14}, $6n$) Es_{99}^{246} become possible. The reaction Cm^{244} (C^{13}, $4n$) No^{253} is believed to have been used in the synthesis of Element 102.

The chemistry of the transuranic elements has become so extensive that we can deal with only a selection of special points. The chief of these from the point of view of inorganic chemistry is undoubtedly the place which these elements are to occupy in the periodic table. This means determining how the electrons are arranged in the various shells and two possibilities seemed at one time to be open. Firstly, they could form a new group of transition elements fitting below the third group (Hf–Au). In this case, the $6d$ electron shell would be built up. The second possibility is that, from the point of view of electron structure, they are analogous to the rare earths (Elements 57–71), in which the fourteen $4f$ electrons are filled in. For the transuranic elements it would be the $5f$ shell which is completed and this is, in fact, what happens.

This conclusion is based both on physical and on chemical evidence. Elements with an incomplete $5f$ shell would show paramagnetism. This is found to be the case, and there is at least a qualitative agreement in this respect between the two groups. The element actinium is believed to have no $5f$ electrons. Spectroscopic evidence, which is difficult to interpret in detail for such heavy elements, shows that there are electrons in the $5f$ shell for uranium and possibly also for thorium. Paramagnetic resonance studies have shown that curium also has these electrons. The series is therefore thought to begin at actinium and the family of elements is called the actinides, by analogy with the lanthanides, which begin with lanthanium, the first of the rare earths. The probable number of f electrons for the two series is shown below.

$Z = 89$	Ac	$5f^0$	57	La	$4f^0$	$Z = 97$	Bk	$5f^8$	65	Tb	$4f^8$
90	Th	$5f^1$	58	Ce	$4f^1$	98	Cf	$5f^9$	66	Dy	$4f^9$
91	Pa	$5f^2$	59	Pr	$4f^2$	99	Es	$5f^{10}$	67	Ho	$4f^{10}$
92	U	$5f^3$	60	Nd	$4f^3$	100	Fm	$5f^{11}$	68	Er	$4f^{11}$
93	Np	$5f^4$	61	Pm	$4f^4$	101	Mv	$5f^{12}$	69	Tm	$4f^{12}$
94	Pu	$5f^5$	62	Sm	$4f^5$	102	No	$5f^{13}$	70	Yb	$4f^{13}$
95	Am	$5f^6$	63	Eu	$4f^6$	103		$5f^{14}$	71	Lu	$4f^{14}$
96	Cm	$5f^7$	64	Gd	$4f^7$						

A number of crystalline compounds of the actinides have been subject to a detailed structural analysis by the methods of X-ray spectroscopy and ion sizes have been deduced. These show a contraction—the actinide contraction—which is exactly analogous to the lanthanide contraction in the rare earth series. The data in Table I illustrate this point for the ter- and quadri-valent actinides which have been examined.

TABLE I

Ionic Radii for Lanthanide and Actinide Elements

	La^{3+}	Ce^{3+}	Pr^{3+}	Nd^{3+}	Pm^{3+}	Sm^{3+}	Eu^{3+}	Gd^{3+}	Tb^{3+}	Dy^{3+}	Ho^{3+}
r in Å	1·06	1·03	1·01	0·99	—	0·96	0·95	0·94	0·92	0·91	0·89
	Ac^{3+}	Th^{3+}	Pa^{3+}	U^{3+}	Np^{3+}	Pu^{3+}	Am^{3+}				
r in Å	1·11	—	—	1·03	1·01	1·00	0·99				
		Th^{4+}	Pa^{4+}	U^{4+}	Np^{4+}	Pu^{4+}	Am^{4+}				
r in Å		0·99	0·96	0·93	0·92	0·90	0·89				

The absorption of light by the majority of ions of the actinides lies in the visible region of the spectrum and is associated with transitions between the shielded $5f$ levels. It is not surprising, therefore, to find that the absorption spectra contain sharp bands similar to those of the lanthanides. Indeed, this was known to be the case for the uranyl ion long before the transuranic elements were discovered.

At first sight we might expect the actinide elements to resemble the lanthanides chemically in forming tervalent ions by the loss of a single $6d$ electron and the pair of $7s$ electrons in the two outermost shells. Here, however, analogy breaks down, for the electrons in the $5f$ shell are less firmly held than are those in the $4f$ and, as a consequence, higher valencies result. This was in fact pointed out as likely to be the case before the new elements were discovered. The oxidation states known at present are shown in Table II. The elements beyond curium are not included in this Table and will be discussed later. Up to and including curium the compounds

of the elements have been made at least in weighable quantities and, for plutonium, in much larger amounts. The oxidation state of + 6 has its maximum stability in uranium and though similar types of compounds are formed up to americium, the

TABLE II

Oxidation States of Lanthanides and Actinides

La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy
					2	2			
3	3	3	3	3	3	3	3	3	3
	4	4						4	

Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf
3	(3)	(3)	3	3	3	3	3	3	3
	4	4	4	4	4	4	4	4	
		5	5	5	5	5			
			6	6	6	6			

lower states gain in stability. The most familiar ion in solution is MO_2^{++}, which occurs in the well-known uranyl salts. The ion UO_2^+ is also known but, for the two lower states, the simple ions M^{4+} and M^{3+} are encountered. The range of solid oxides formed by uranium and the three elements which follow it is shown in Table III.

TABLE III

		Pu_2O_3	Am_2O_3
		Pu_4O_7	
UO_2	NpO_2	PuO_2	AmO_2
U_3O_8	Np_3O_8		
UO_3			

Table IV, similarly, shows the types of halide formed by these four elements (X = F, Cl, Br, I).

TABLE IV

UF_6, UCl_6	NpF_6	PuF_6	
UF_5, UCl_5			
U_2F_9			
UX_4	NpF_4, $NpCl_4$	PuF_4	
UX_3	NpX_3	PuX_3	AmX_3

Many new techniques have been developed during the exploration of this field. It has involved work with radioactive species of all sorts and much attention has been devoted to the elimination of all risk arising from the exposure of workers to radiations. This has ranged from the shielding of complete chemical installations in the plutonium industry to devices for remote control in laboratory work. Where there is dust hazard dry-boxes have also been widely used. All sorts of microchemical techniques have also been adapted for work with minute amounts of radioactive substances.

Special mention must be made here of two special techniques which have contributed greatly to the study of the whole field of transuranic element chemistry.

These are solvent extraction and the use of ion-exchange resins. It has long been known that certain inorganic compounds are soluble in organic solvents and that extraction from an aqueous phase into a non-aqueous phase is possible. Uranyl nitrate will, under suitable conditions, pass from water into ether and, by changing the conditions, may be made to revert to an aqueous phase. This is the basic operation in preparing pure uranium compounds for atomic energy purposes, for it makes possible the removal in one stage of impurities such as rare earth ions which, if present in uranium, may actually stop the normal chain reaction. Other solvents (e.g. higher ethers, tributyl phosphate and methyl isobutyl ketone) act similarly. Solvent extraction is also used in the separation of plutonium from uranium and fission products with which it is mixed when uranium rods are withdrawn from the nuclear reactor.

Compounds with ion-exchange properties may be inorganic or organic. To the former class belong the natural and synthetic zeolites. The organic ion exchangers are synthetic resins of suitable type with free acidic groups (e.g. $-SO_3H$, $-CO_2H$ or phenolic –OH) built into their structures. Such compounds can absorb cations and are known as cation-exchange resins. The other type, called anion-exchange resins, has basic groups such as $-NH_2$ in the structure. In this case, anions are taken up. We will confine ourselves to a brief consideration of the use of cation-exchange resins. When such a resin is washed with a dilute acid it is converted into the so-called acid form, in which all or most of the cation places are occupied by hydrogen ion. When a solution containing several cations is passed through a column of the resin in its acid form the cations are taken up and hydrogen ions released. The cations are, however, held with varying degrees of firmness, depending on their size, charge and basicity, and when, for example, a complexing reagent is passed through the column they are released in a definite sequence, with a degree of separation of the constituents of the original mixture which naturally varies from case to case. This is one of the best ways of separating the rare earth ions, which are so alike chemically as to make normal separation procedures very tedious. The same method has been used to separate the heavier transuranic elements from one another and from rare earth fission products with which they are often associated. Each ion is eluted in turn and gives a more or less well-defined peak in the curve obtained by plotting the radioactivity of the effluent liquid against the drop number. This is illustrated for a typical series of radioisotopes of the rare earths and for a series of transuranic elements in the diagram below.

The elements of highest atomic number are eluted first in both series and it is readily seen that, for a new element, the time at which it will emerge from the column can be predicted. The radioactive characteristics to be expected in the new isotope may also be predicted so that identification of a new element is possible even if, as in the case of nobelium, very few atoms of it are present.

The chemistry of the elements beyond curium has been established entirely by tracer techniques. In the case of californium, for example, the elution properties of the ion show that it is present initially in the tervalent state. When a solution containing californium is treated with very strong oxidising agents it is found not to be coprecipitated with zirconium phosphate, which would carry down with it Cf^{4+}. The inference is that oxidation does not produce this higher oxidation state. The formation of a solid compound with a higher oxidation state is not excluded. It is probable that in time, work on a macroscopic scale for solid compounds and elements up to einsteinium will become possible.

When we attempt to look forward in this field the first question which comes to mind is whether there are more transuranic elements to be discovered. That any could occur naturally is out of the question. There are limits to the natural stability of atomic nuclei and all the isotopes of elements beyond uranium are so short-lived that they can only occur on the earth if they are continually generated. Not only

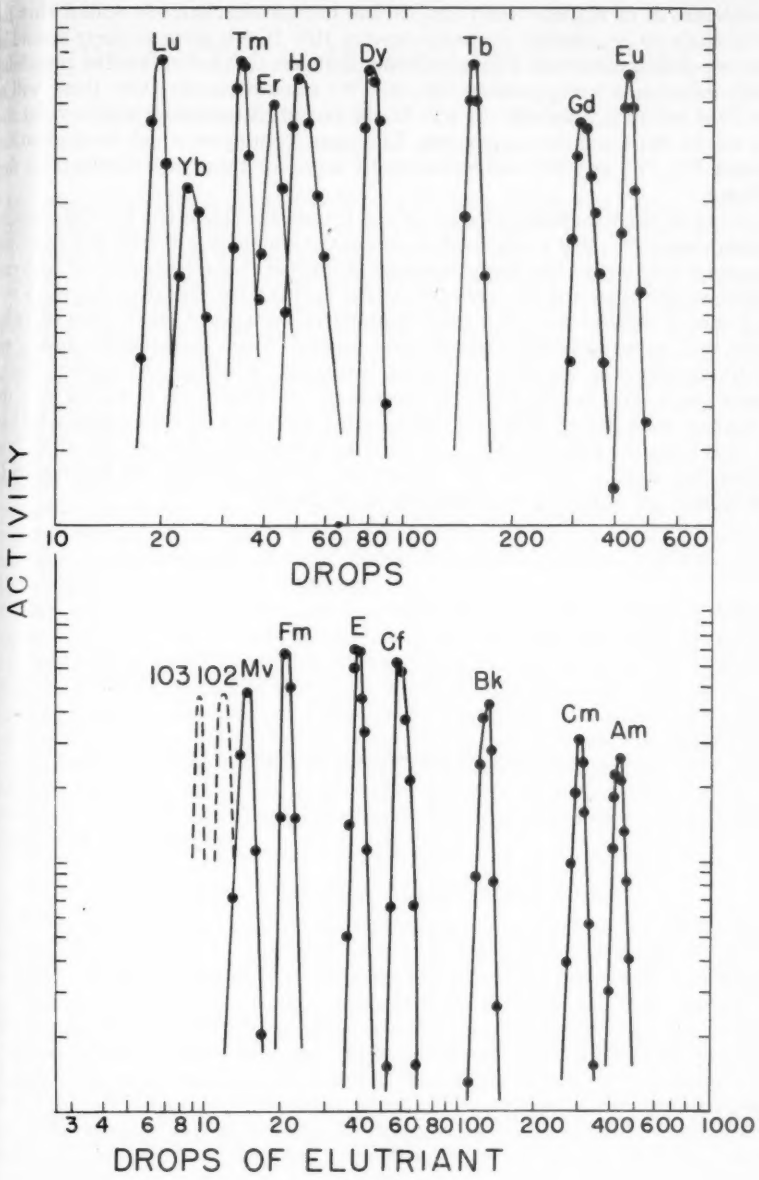


Fig. 2

(Reproduced by courtesy of Methuen & Co., Ltd., from *The Chemistry of the Actinide Elements* by Seaborg and Katz.)

radioactive decay a source of instability but spontaneous fission, the most miliar instance of which is that of U^{238}, becomes increasingly important as a mitation on the survival of nuclei as these become very heavy. Spontaneous ssion is not directly related to nuclear mass, but a great deal is now known about

the systematics of nuclear decay and it has been authoritatively stated that the limit is likely to be reached at about element 108. If this were so there would be six more possible elements. I have no doubt that research to explore the possibility of their isolation is being pressed forward. We must remember that these will be short-lived isotopes. Element 103 will be the last of the actinides and beyond that there will be more transition elements. The same techniques which have identified elements 100, 101 and 102 will undoubtedly serve for these new elements as they are found.

Turning to the future significance of the transuranic elements for the chemist, it appears inevitable that a new field of chemical technology will arise in connection with power production and the processing of nuclear fuel elements. Our universities cannot, as a rule, hope to have the special facilities for experimental work with high levels of radioactivity, but basic training in the special techniques of radioactivity will certainly become much more general. More specialised training will have to be given in specially equipped institutes or within the atomic energy industry itself. This is a field which abounds in opportunity. It is a good sign that our exciting atom power stations are run for the most part by young men who have been in it from the beginning and we need have little fear that the scientific challenge of this work, as well as the opportunity which it offers for service to the country, will fail to attract our graduates of the future.

BENEATH THE EARTH'S CRUST[1]

ADDRESS BY PROFESSOR L. R. WAGER, F.R.S.
PRESIDENT OF SECTION C

PLANS FOR DRILLING a deep bore hole into the rocks below the Earth's crust were made at the recent meeting of the International Union of Geodesy and Geophysics in Toronto. It is hoped to obtain samples of the material of the upper mantle and to get direct information on the temperature and other physical conditions. A bore hole some ten kilometres deep is aimed at; this is twice the depth of any so far drilled but, if sufficient money is also forthcoming, it is believed to be not beyond present engineering skill. It is interesting to find an early situation roughly repeating itself. Darwin's hypothesis on the origin of coral reefs could be tested, it was realised, by boring deeply into an atoll and Darwin himself, in a letter to Agassiz in 1881, wrote: 'I wish some doubly rich millionaire would take it into his head to have borings made in some of the Pacific and Indian atolls and bring home cores for slicing from a depth of 500 or 600 feet. . . . If I am wrong (in my hypothesis) the sooner I am knocked on the head and annihilated the better.' This is what is wanted today, only we are more ambitious as to depth. It is wanted for essentially the same reason: to knock on the head any hypotheses about the nature of the mantle which are found wanting.

In the near future, if this ambitious plan succeeds, we shall have available samples of the rock from below the earth's crust and I have therefore thought that the present is a good time to review the existing ideas on the nature of the material which the boring should find and to consider the significance to petrology if this material proves to be similar to the alpine type peridotites. As a petrologist I have found such a review a difficult task because of the particularly rapid advances which are being made in geophysics and geochemistry at the present time. However, I have persisted in the course set because I am sure that the nature of the subcrustal material is of crucial importance for several major aspects of geology, for instance the origin of basalt magma.

A CURRENT PHYSICAL MODEL OF THE EARTH

Before considering what lies beneath the crust it is necessary to say what is meant nowadays by the earth's crust and to summarise present ideas on the general framework of the earth provided by physical methods of investigation. The term crust had an obvious meaning when it was believed that the solid outer layers of the earth rested on a liquid interior from which came the molten rock extruded by volcanoes. However, from the small amplitude of solid earth tides and from seismic evidence it is clear that the outer half of the earth, at any rate, is solid—Lord Kelvin, the great Glasgow physicist, likened the strength of the earth to that of steel. The core at depths beyond 2,900 km is apparently partly or completely liquid, but the core is beyond my immediate interests. It will be some time yet before even the most optimistic plan a boring to sample the earth's core. The term crust now usually means the rocks down to the Mohorovičić discontinuity, a rather well-defined break in physical properties occurring at a depth varying from 10 to 60 km, which is mapped out, as it were, by determination of the variations in earthquake wave velocities. In the rocks above the Mohorovičić discontinuity, the earthquake wave velocities increase downwards but differently in different places, the inference being that the rocks are variable in composition, as is indeed to be

[1] Address delivered on September 1, 1958, at the Glasgow Meeting of the British Association.

expected by reasonable extrapolation of the surface geological evidence. On the other hand below the Mohorovičić discontinuity, that is in the uppermost part of the mantle, the velocity of compressional earthquake waves is on the whole surprisingly constant at about 8 km per sec, implying that the subcrustal rock is broadly of uniform composition. Some lack of uniformity below the Mohorovičić discontinuity is however to be expected but on a relatively much smaller scale than that in the crust and there is seismic evidence indicating this, as Professor Hawkes has stressed in a recent address to the Geological Society.

During the last decade seismic exploration of the central Pacific region has proved the existence of a relatively simple structure. According to Hill (1958) certain layers with characteristic compressional wave velocities and fairly constant thicknesses are present at many widely separated positions. It is considered that the following succession represents the structure of the outer earth in the central Pacific region and that a fairly similar structure is characteristic of parts of the other oceanic areas.

Generalised Oceanic Column

	Thickness in km	*Vp* in km per sec	Sp. gr. (after Hess)
Sea water	4·5	1·5	1·03
Unconsolidated sediments .	0·45	2	2·3
Intermediate layers . .	6·4	4–6·7	2·9
Mohorovičić discontinuity			
Subcrustal material . .	—	8·1	3·3

In continental areas there is more variability. Below the obviously complex sedimentary and igneous upper part there are usually rocks giving compressional wave velocities from about 5–6·5 km per sec but there is lateral variability and the increase with depth is also less regular. The Mohorovičić discontinuity may vary from a depth of 20 to 50 km, but at whatever depth it is found, the compressional wave velocity in the sub-Mohorovičić material is about 8.

Generalised Continental Column

	Average thickness in km	*Vp* in km per sec	Sp. gr. (after Hess)
Variable continental rock (sial)	35	*c.* 5–6·5	2·7–2·9
Mohorovičić discontinuity			
Subcrustal material . .	—	*c.* 8	3·3

Gravity surveys over the continents and oceans have shown that a general hydrostatic balance prevails within the earth. Therefore the densities of the various layers, the thicknesses of which have been defined seismically, must be such as to give equal masses in oceanic and continental columns. In the above tables are given likely figures for the densities of the layers according to Hess (1954 and 1955). This is the kind of upper earth structure which I am going to assume in reviewing the problem of the nature of the upper mantle.

The probable temperatures within the outer layers of the earth are also significant for any enquiry into the nature of the subcrustal material. Simple extrapolation downwards of the temperature gradient established by measurements in the outermost skin of the earth soon fail to carry conviction because of doubt about the amounts of the naturally radioactive elements which produce heat by disintegration, and doubts whether or not convection occurs in the solid mantle. Recent observations at sea (cf. Bullard, Maxwell and Revelle, 1956), have produced

the unexpected result that the heat loss through the Pacific floor is, on the average, the same as through the continental crust, despite the much greater content of heat-producing radioactive elements in crustal rocks. Although there are great differences in the various hypotheses about the temperature at great depths in the earth, there is general agreement that the material down to the core is crystalline and therefore that the temperatures are everywhere below the melting point of the material at the pressure existing. A recent review by Verhoogen (1956) has shown general agreement that the temperature in the uppermost part of the mantle at 50 km depth is about 700° C., which is well below the melting point of the likely material. If independent evidence succeeds in providing knowledge about the composition of the mantle then laboratory determinations of the melting point of such materials under high pressure would give a *terminus ad quem* for the temperatures. This is a typical case of the interdependence of the compositional and physical knowledge about the deep earth; if we only knew the one, then the other could be deduced.

Fitting Substance to the Physical Model

What substance can be fitted to the physical earth model which I have sketchily indicated? From the detailed surface studies of orogenic belts and basement complexes we can extrapolate downwards to get some idea of the composition of the upper part of the continents, but I confess to feeling quite uncertain how far such extrapolation can take us. In basement complexes and the roots of mountains the detailed picture is wildly complex; we must hope that with increasing depth there comes increasing simplicity, otherwise our chief physical tool, the seismic method, may prove too insensitive. Although the material forming the continental crust must be very heterogeneous, its average composition is probably roughly granodioritic and Suess gave it the name sal (later changed to sial by Wegener) because of its presumed richness in silica and alumina. This is a term which still has a general usefulness.

Below the depth of extrapolation from conventional geological mapping, and over the 71 per cent of the earth which is covered by water, the most important messengers bringing direct knowledge of the chemical composition of the depths of the earth are the igneous rocks and the related volcanic gases, thermal waters and mineral deposits. The great diversity of igneous rocks has tended to obscure some of their broad relationships. Daly did much to get true perspective by estimating the relative amounts of the chief types and this led him to the clear realisation of the abundance and petrogenetic significance of basalt. Simultaneously, from experimental work on silicate melts Bowen showed how fractional crystallisation of basalt magma could give rise to many rock types. The widespread occurrence of basalt in both continental and oceanic areas shows that over the whole earth a layer is present from which by some means, basalt magma can be generated. Daly developed the hypothesis that beneath the continental and oceanic crust there was uncrystallised basalt, a basalt glass, which behaved as a solid in relation to earthquake waves but which flowed under slight shearing stresses if they were applied for a long time. The vitreous layer postulated by Daly explained the delicate response to the loading and unloading of the crust by ice sheets and the flow of material needed to account for isostatic adjustments during the erosion of mountain ranges. Such a glassy basalt layer could also provide fluid magma of its own composition simply by reducing the hydrostatic pressure upon it. Daly's hypothesis, attractive as it was to many geologists, has had to be abandoned in the face of experimental work by Birch and others on the physical properties of materials under high pressure and temperature. Reasonable extrapolation from laboratory data indicates that basalt glass, at the temperatures expected at 10 or 20 km within the earth, would not give the observed earthquake wave velocities and, with

certain reservations, Daly (1946) subsequently accepted this view. The facts of isostasy show, however, the necessity for some layer within the earth which is capable of flow under slight but protracted shearing stress. This condition, considered so important by Daly, is still a requirement for any earth model.

The material below the crust is such that compressional earthquake waves travel in it with a velocity of about 8 km per sec. In the search for likely mineral assemblages which would have the necessary physical properties to produce this velocity it was soon found that the olivine-rich rocks, peridotite and dunite, were the only likely ones among the rocks accessible to us. More extensive observation at high pressure and relatively high temperatures by Birch and others has strengthened the early view that the subcrustal material is likely to be largely olivine, but it would be of interest to carry out more detailed experimental investigations. The ultrabasic rocks are also reasonable as sub-Mohorovičić discontinuity material because of their higher density and higher range of melting temperature compared with the gabbros, diorites and granites believed to form the overlying crust.

In his classic book on the evolution of igneous rocks Bowen made an important contribution to the problem of the nature of the material at moderate depths in the earth from which basalt might be expected to be derived. In the early chapters of the book he gave much evidence, based on silicate chemistry combined with field data, that fractional crystallisation of basalt magma could yield a variety of igneous rocks. In a later chapter he considered essentially the reverse of this process, that is the fractional melting of rocks, and he showed the plausibility of basalt magma being produced by partial melting of peridotite. Historically, Bowen's suggestion, to which I shall return later, was contemporaneous with Daly's advocacy of the hypothesis of a vitreous basalt shell. At that time petrologists were, for the most part, not concerning themselves with the ultimate origin of basalt but only with its various differentiates and the mechanism by which one basalt magma type might be derived from another. They tended to accept Daly's hypothesis as being an adequate way of getting the basalt magma in the first place and Bowen's suggestion, until recently, remained undeveloped.

One of the early facts established about the earth as a whole concerned its density. Newton made a celebrated estimate that the density of the earth would be five or six times that of water and experiments on the mountain Schiehallion, in Perthshire, provided the first experimentally determined value. It was early realised that the high density of the earth, now accepted as about 5·4, would not be likely to result from compression of rocks common at the earth's surface and this has been amply confirmed by the experimental work of Bridgman. In 1863 the great French experimental geologist Daubrée suggested that the high density of the earth might be due to a nickel-iron core similar in composition to the iron meteorites, and that the overlying material, now called the mantle, might have the composition of the stony meteorites. This is the beginning of an hypothesis which is nowadays basic for much geochemical argument. The first seismic evidence for a sharp discontinuity about halfway to the earth's centre was given by Oldham (1906) in a paper to the Geological Society of London. The discontinuity was interpreted as the result of a liquid, nickel-iron core underlying a silicate shell. A decade ago, the possibility of the core being undifferentiated silicate material, in a metallic state as a result of the very high pressure, was much discussed by Kuhn, Rittmann (1941) and others, but the consensus of opinion has gone back again to Daubrée's idea that the core is essentially of nickel-iron composition.

If meteorites come from a disrupted planet like the earth they should provide indirect evidence of the overall composition of the earth. Thus in developing hypotheses about the constitution of the earth one method of proceeding is to assume an overall composition like that of average meteorite material falling on

the earth (one part of average stony meteorite to 43 parts of average chondritic meteorite) and then from present chemical and thermodynamic knowledge to fit this composition in the most plausible way into the framework of the earth deduced from physical measurements. For many purposes it has become usual to equate the composition of the outer silicate part of the earth, the mantle and the crust, with the composition of the stony meteorites. I shall not review the implications of this hypothesis as it would take me too far into geochemistry and moreover, it does not seem to be applicable without qualification. No doubt broadly speaking the silicate part of the earth consists mainly of magnesium, iron, silicon and oxygen, as suggested by analogy with the composition of chondrites, but we shall show later that, without qualification, the simple hypothesis does not fit with the mineralogical and chemical composition of basalt, which I believe to be the most abundant derivative of the mantle. The processes in the disrupted planet from which the meteorites have come do not seem to have been the same as those producing the various earth shells.

The Composition of the Upper Mantle

Various lines of evidence predisposed geologists to think of peridotite, or some related olivine-rich rock, as forming the material beneath the crust and I now wish to consider this in more detail. I shall begin by enquiring whether any rock now accessible to us can be accepted as an actual sample of the material from below the Mohorovičić discontinuity. Professor Hess would answer this question definitely in the affirmative as he assigns certain peridotites and serpentinites to a deep, subcrustal source (1955). He points out that the accessible ultrabasic rocks belong to two groups; one has a systematic connection with gabbros and the other has not. The latter group which, following Benson, are usually called the alpine type of ultrabasics, are considered by Hess as having once formed part of the earth's mantle.

As an example of peridotites associated with gabbros we cannot do better than consider for a moment the beautiful peridotite and dunite of Sgurr Dubh in the Isle of Skye which is associated with the Cuillin gabbro. Harker thought of these peridotites as representing a separate, early intrusion; Bowen later gave reasons for believing that they, and other peridotites, are the result of accumulation of early olivine crystals from a large volume of olivine gabbro magma. Despite great geological complexity, and even greater annual rainfall, remapping by Oxford research students is gradually building up detailed evidence that the Skye peridotites, eucrites and gabbros are the result of bottom accumulation of crystals to form a layered complex. In the Rhum, Stillwater and Duluth layered intrusions there are peridotites and related rocks to which a similar origin is ascribed. This group of olivine-rich ultrabasics may be conveniently distinguished as autochthonous cumulites, that is rocks formed by the accumulation of early crystals which have not been much moved from their place of origin.

The other group of ultrabasic rocks, as Benson (1926) first showed, tend to form belts of discontinuous masses in fold mountains or island arcs. They are not found associated with gabbro except, it is believed, accidentally; thus without some ancillary hypotheses they cannot be regarded as autochthonous cumulites from basic magma. In many cases the alpine ultrabasic masses are entirely serpentinised; in others, relics of olivine and enstatite are preserved, while in a few cases, such as at Dun Mountain in New Zealand and in New Caledonia, there are fresh dunites and harzburgites. In the alpine serpentinites, if they have not undergone too much mechanical shearing, mesh and bastite structures are present indicating derivation from the olivine or enstatite of peridotites which were essentially the same as the rarer, fresh examples of alpine type ultrabasics. From this evidence and also from Bowen's and Tuttle's experimental work on the system MgO-SiO_2-H_2O (1949),

it seems certain that most alpine type serpentinites are derived by alteration from peridotites.

In 1937 Hess made the bold suggestion that the alpine type ultrabasics represent the material of the mantle pushed into the crust when it formed a down buckle, or tectogene. Meinesz had previously suggested that such a tectogene underlies the belt of negative gravity anomalies in the East Indies, and that tectogenes may be temporary features of many fold mountains. The evidence offered by Hess for his suggestion was first, the peculiar geographical distribution of the ultrabasic masses in two zones on either side of the main orogenic axes and secondly, the fact, already mentioned, that the alpine ultrabasic rocks are not associated with the large quantities of gabbro which would be necessary if they were derived by accumulation of olivine from a parental basic magma. These ultrabasic masses are perhaps extraordinary enough to have had the origin suggested by Hess and it is of the greatest importance for geologists to decide whether or not the alpine ultrabasics are actual samples of the mantle.

As Bowen first pointed out, if ultrabasic rocks were formed from a melt of their own composition, intruded into the position in which they are found, they would have produced very high grade thermal metamorphism because of the high temperature of a liquid of such a composition. Characteristically, however, the metamorphism at the contacts of the alpine type ultrabasics is even less than that at gabbro contacts. Injection of high temperature magma to give the alpine type serpentinites and peridotites is thus ruled out by field evidence. At one time Hess (1939) suggested that the serpentinites were formed from ultrabasic magma rich in water—a sort of low temperature hydrothermal magma—but the experimental work by Bowen and Tuttle (1949) showed that liquid of such a composition does not exist but instead crystals of olivine coexist with water vapour down to a temperature of 400° C., when serpentine replaces the olivine. The mechanism for the emplacement of the alpine peridotites, suggested by Bowen and Tuttle, is the same as that which Bowen had previously suggested for the peridotite dykes of Skye, namely intrusion, at moderate temperatures, of masses of olivine crystals lubricated by water vapour. No convincing evidence of how the alpine peridotites and their serpentine derivatives were emplaced has yet been found; as Hess (1955) says, some vital piece of information is still missing. If olivine-rich mantle material can be in such a physical condition that at times it is injected into the crust as a relatively low temperature, quasi-solid mass, it is likely that autochthonous olivine cumulites would also pass through a stage when, under suitable tectonic conditions, they could similarly move into new positions and this is, indeed, one of the suggested modes of origin for the peridotite dykes of Skye.

There is still another distinct type of occurrence of the ultrabasic rocks which is sufficiently peculiar to lead to the suggestion that they may be pieces of subcrustal material. I refer to the so-called 'olivine nodules' found in certain basalt lavas, and as cored bombs in certain basaltic pyroclasts. These nodules are characteristically about the size of a closed fist; they are light green in colour and consist of olivine-rich ultrabasic rock. So striking are they that as long ago as 1787 Le Lièvre, who found in the Pyrenees the rock later to be named lherzolite, suggested that the 'nodules à olivine' already described by Faujas de Saint Fond from the lavas of certain Quaternary volcanoes of central France, were derived from the break-up of a lherzolitic rock in the volcanic basement. The distribution of the olivine nodules is oddly irregular; for instance many Tertiary and Quaternary basalts in Germany and France contain them, but they are unknown from British Tertiary basalts where they would surely have been found if they exist. In mineral and chemical composition the olivine nodules are frequently exactly like alpine type dunites, such as that from Dun Mountain; at other times the nodules consist of olivine and enstatite with textures resembling the alpine type harzburgites. On

the other hand in Hawaii, associated with typical dunite nodules, are others having some of the textural features of the cumulites of layered intrusions, in particular poikilitic plagioclase or augite surrounding more or less idiomorphic olivine crystals. The close similarity between typical olivine nodules and alpine type dunites has been confirmed recently by Ross, Foster and Myers (1954) who have made a thorough chemical investigation of their constituent minerals. To take one example, they have shown by analysis that the composition of all the olivines of the nine olivine nodules investigated from such widely separated localities as Germany, California, Hawaii and Korea, range in composition only from $Fo_{89\cdot7}$ to $Fo_{91\cdot0}$, the olivines of the four alpine type dunites investigated being between $Fo_{91\cdot4}$ and $Fo_{92\cdot1}$. The enstatites and chromites from both sources are equally similar. Whether the slight differences found by these investigations will prove to be significant is an open question; at the extreme magnesium-rich end of the olivine solid solution series differences of one molecular unit may perhaps turn out to be of importance. It was shown by Hess (1939) that the olivine of peridotites associated with gabbros, here called the autochthonous olivine cumulites, was more iron-rich on the average (Fo_{86}), than those from alpine type peridotites (Fo_{90}). While it is likely that the olivines of the autochthonous cumulites are usually the more iron-rich it is apparent that some are not. Thus the autochthonous cumulites of Rhum, particularly the harrisites, may contain olivine which is Fo_{89}, and olivine from an autochthonous dunite forming part of a small layered mass near the Skaergaard intrusion, East Greenland, is Fo_{92}.

As there is an overlap in the composition of olivines from alpine type dunites, olivine nodules of basalts and autochthonous dunites, mineral composition cannot by itself be used to distinguish between possible subcrustal peridotite material and autochthonous layered material formed as a cumulite from basic magma intruded into the crust. An independent piece of evidence on the source of the olivine nodules in basalts comes from the recent determination of the isotopic composition of the very small amounts of lead which can be extracted from these rocks (Davis, Aldrich *et al.*, 1956). The isotopic ratios are those of 'young' lead and suggest that the particular olivine nodules investigated were formed as a relatively recent differentiate of basalt magma rather than being fragments of the original mantle of the earth.

As a petrologist confronted on every hand with basalt and its differentiates, I find the most compelling argument for the subcrustal material being peridotite resembling the alpine type dunites and harzburgites, to be that such material would provide, by partial melting, a source of basalt magma. A magnesium-rich olivine (Fo_{90}) begins to melt in the laboratory at ordinary pressures to give a more iron-rich liquid together with some more magnesium-rich olivine (Fo_{91}, Fo_{92}, etc., according to the amount of melting). The liquid, if separated from the crystal mush, would on crystallisation produce a first crop of olivine crystals of about Fo_{90} changing as fractionation proceeded to less magnesium-rich olivine and thus corresponding to the usual composition of early olivines in many basalt lavas. If the upper mantle were forsteritic olivine, about Fo_{90}, together with small amounts of other constituents as in the alpine peridotites, then partial melting could provide a liquid of general olivine basalt composition. It is of little value at the moment to consider the remelting process in detail. Nevertheless in the table below is presented a weighted average of the only four good analyses of non-serpentinised alpine type ultrabasics (there are all too few analyses of material with less than 2 per cent H_2O) and from this has been subtracted 80 per cent of olivine of composition Fo_{94}. The remaining 20 per cent is approximately the composition of the liquid which would be produced by 20 per cent of partial melting and it has some resemblance to the composition of basalt magma. Actually no closer agreement would necessarily be expected because the peridotites, the analyses of which have been used in making the estimate, may be of mantle material from which

basalt has already been abstracted by partial melting and, if so, further melting obviously could not give typical basalt liquid.

Rough as the compositional argument is at present, I believe it to be the best available evidence for the nature of subcrustal material. Basalts, being what they are, must have been derived from something approximating to the composition of alpine type peridotite if they were produced by partial melting, and the physical conditions within the earth seem to compel us to the view that partial melting of material at considerable depth is the most likely source of basalt magma. Unfortunately this argument neither proves nor disproves Hess's view that alpine peridotites are actual mantle material, because an early autochthonous cumulite, formed

TABLE I

Averages of four dunites (I), of chondritic meteorites (II), and an estimate of the composition of the liquid produced by 20 per cent of melting of average dunite (III)

	I*	II†	III‡
SiO_2	41·9	47·0	48·6
Al_2O_3	1·1	3·1	6·2
Fe_2O_3	0·9	—	5·1
FeO	6·5	15·4	9·6
MgO	46·2	29·5	23·7
CaO	1·1	2·4	6·2
Na_2O	0·1	1·2	0·6
K_2O	0·005	0·2	0·03
H_2O^+	0·4	—	
H_2O^-	0·1	—	
TiO_2	0·1	0·14	
P_2O_5	0·01	0·26	
MnO	0·1	0·3	
Cr_2O_3	1·4	0·45	
NiO	0·3	—	
$\frac{Fe \times 100}{Fe + Mg}$	7·4	22·8	

* Weighted average of four dunites: Dunite (× 3), New Caledonia (Lacroix, 1942, p. 18, No. 2); Saxonitic dunite, St. Paul's Rock (Tilley, 1947, p. 490, No. 1); Peridotite, Sweden (du Rietz, 1935, p. 143); Olivine nodule, California (Hess, 1955B, p. 426, *b*).

† Average of silicate phase of 94 chondrites (Urey and Craig, 1953, p. 55).

‡ Composition obtained by removing 80 per cent of olivine, Fo_{94}, from average peridotite (column I). This approximates to the composition of the liquid which would be produced by 20 per cent of partial melting.

from basalt magma which was itself produced by partial melting, would have an essentially similar composition.

The hypothesis that the upper mantle has the composition of alpine dunite and harzburgite should be considered in relation to the view that the mantle has the overall composition of chondritic meteorites. A new average for the composition of chondritic meteorites has recently been provided by Urey (1953). In this the ratio of iron to iron plus magnesium is 0·228, a value very different from the 0·074 of the average peridotite just considered. Partial melting of average chondritic material could not provide a sufficiently magnesium-rich liquid and it is unlikely, although perhaps not impossible, that crystal fractionation of average

chondritic material as discussed in the next section, could produce an upper mantle from which such a magnesium-rich liquid might come. To get over these difficulties Urey (1953) and now Ringwood (in the press) have suggested that the behaviour of iron and silica has been somewhat different in the earth from what it was in the parent meteorite body. Thus the connection between average chondrite composition and the composition of the mantle becomes hedged about with qualifications. The composition of the earth's mantle will have to be decided ultimately on direct terrestrial data and among these the composition, and proper understanding of the origin of basalts will perhaps be the most important.

The Formation of the Earth's Silicate Shell—Analogies with Layered Intrusions

I now turn to consideration of a mechanism for the formation of the earth's silicate shell, based largely on analogy with layered intrusions. As I prepared this section I frequently thought of the final words of the presidential address prepared by Professor H. H. Read for the Dundee meeting, a characteristic feast but unfortunately never delivered because of the declaration of war. He wrote, '. . . just as things too absurd to be said can yet with perfect propriety be sung, so views too tenuous, unsubstantial and generalised for ordinary scientific papers can yet appear with some measure of dignity in presidential addresses'. A less encouraging attitude is that of Whewell, who in his history of the Inductive Sciences said of some theories that they were 'premature attempts to snatch them before their time'.

From physical measurements various earth models can be constructed to provide consistent, possible pictures of the earth; the models are then tested against new facts as they are obtained. Similarly from chemical and petrological data compositional models can be built up. Most of our earth models combine these two approaches but for the moment I wish to concentrate on the compositional aspect. Even if the earth were originally a cold accumulation of material it is, nevertheless, usually considered to have passed later through a stage of being a hot liquid. The manner of crystallisation of such a liquid, as Buddington has suggested (1943), may have been comparable with that of layered intrusions. In these intrusions, of which I believe the most instructive example is still the Skaergaard of East Greenland (Wager and Deer, 1939), high temperature minerals, as they crystallised from the magma, accumulated at the bottom of the liquid to build up successive layers of igneous rock. Although cooling mainly took place from the top surface of the liquid, the crystals separated out at, or near, the bottom because a convective circulation was established, which produced an average thermal gradient less than the increase in melting point of the minerals with the increase of pressure due to descent in the liquid. Lord Kelvin considered the likely sequence of events during the solidification of a molten globe. He postulated that due to surface cooling a solid crust formed which broke up and, being heavier than the liquid, foundered. He considered that to begin with the foundered crust would be remelted in depth but that later, due to general cooling of the whole earth, the foundered blocks would remain unmelted. From then on a central core consisting of a honeycomb mass of sunken blocks and liquid developed and grew until the whole earth was in this state. Later Kelvin (1897, p. 14) seems to have realised the possibility of bottom separation of crystal phases, a possibility made much clearer by the data and arguments put forward by L. H. Adams (1924). Jeffreys, in the second edition of his great book, *The Earth* (1929), accepted the view that convection currents would exist in a molten earth and that, as a consequence, solidification would take place from the bottom upwards. He suggested that such bottom accumulation might be so rapid that perhaps crystallisation of the whole silicate part of the earth

happened in less than a million years (Jeffreys, 1952, p. 269), although possible factors which could not be evaluated might have considerably extended the time.

Any crystal accumulate, at the time of its formation, is surrounded by liquid. From simple experiments with lead shot and other pieces of metal of various shapes immersed in water, the amount of liquid relative to crystals may be seen to lie between 40 and 60 per cent of the whole. In the case of the layered intrusions so far studied the behaviour of the interstitial liquid during further cooling is variable and far from obvious; in the development of the earth's mantle it is likely to be even more complicated and at the same time of especial significance.

In the Skaergaard layered intrusion the bulk of the liquid trapped between the crystals of the primary precipitate, crystallised out slowly over a considerable range of temperature. There was little reaction between the liquid and the early formed crystals to establish equilibrium between them at the lower temperatures; instead the interprecipitate liquid mainly formed lower-temperature solid solution zones round the original crystals. In the Rhum layered intrusion described by Brown (1956) and perhaps also in the Stillwater layered intrusion being described by Hess (in the press), the primary precipitate crystals grew by diffusion of the required substance from the overlying liquid while they lay within a few inches of the temporary top of the crystal pile. The extent to which the diffusion effect operates depends on the rate of accumulation of the crystal precipitate and on other factors such as the ease of nucleation of the crystals. Growth of the primary precipitate crystals as they lie at the bottom of the liquid requires removal of heat. In the silicate shell of the earth, as in large intrusions, the heat loss must have been through the overlying liquid. For this to take place without producing a new shower of crystals from the liquid, I suggest that there must have been delay in crystal nucleus formation resulting in some degree of supersaturation. At the time of the crystal accumulation to form the mantle, various processes may have reduced the amount of the pore liquid but it is unlikely that they would succeed in eliminating it altogether. The original mantle of the earth should consist, therefore, of the crystal phases appropriate to the temperatures and pressures obtaining together with a certain amount of the contemporaneous liquid.

The composition of any pore liquid trapped between the primary precipitate crystals will have the composition of the liquid phase existing at that time and will be different from that of the contemporary crystal precipitate. Crystallisation of the pore liquid will take place over a range of temperature as cooling slowly proceeds; in the case of the mantle the time required may be measured in millions of years and indeed it may not yet be complete because of the high temperature still obtaining. In the case of the accumulate forming the mantle there would probably be time enough for the original precipitate to be made over by reaction with the liquid to solid solutions in equilibrium at the lower temperatures.

While the crystal phases separating as an accumulate will contain only such chemical elements as can enter the particular crystal structures, the trapped interprecipitate liquid contains all the elements of the contemporary liquid phase. Among these will be the bulk of the naturally radioactive elements such as uranium, thorium and potassium and thus the amount of pore liquid becomes a vital factor in the heat balance of the crust.

By analogy with layered intrusions we may expect to understand the behaviour of a bottom accumulation of crystals up to a thickness of a few kilometres. If, however, the pile of loose crystals, surrounded by a liquid of lower specific gravity, is a hundred or so kilometres thick, the excess of weight of the column of crystals over the column of liquid will become an important factor in the situation. In the deeper part of the column the crystals would be expected to go into solution where in contact with each other and to reprecipitate elsewhere, so that the rock should develop a metamorphic texture. Any disintegration of the crystals, due to higher

pressures in the crystal network than in the surrounding liquid, should produce a tendency for the upward migration of the liquid relative to the precipitate. These two processes only emerge as significant in thick crystal accumulates and they have not been certainly shown to occur in the known layered intrusions. They should, however, be operative and highly important in the mantle, except perhaps in the uppermost part. The extent to which the original interprecipitate liquid of the mantle has been eliminated may be decided, one day, from the heat balance of the earth because the heat producing radioactive elements must originally have been present mainly in the interprecipitate liquid. The postulated interprecipitate part of the mantle is just what is required in Wilson's hypothesis (1957) on the production of the earth's crust by the expulsion from the mantle of the part having lower temperatures of crystallisation, a process which he envisages as taking place continuously throughout the whole of geological time.

The question of the actual crystal phases separating in the mantle as a bottom accumulate cannot be decided by direct observation at present, nor is it within the scope of present experimental techniques. We are, however, on the threshold of important advances, for pressures corresponding to a depth of 250 km within the earth are being attained in the laboratory under fairly well controlled conditions. All that can be said at present is that the crystal phases separating in the lower mantle, including perhaps a cubic modification of olivine, as first suggested by Bernal (1936), are such as to lead to a liquid capable of giving rise to the upper mantle and crust.

In speculating on the later stages of fractionation of the silicate liquid to give the upper part of the mantle and crust, we shall make use of the analogy with layered intrusions and shall also assume that the alpine type of ultrabasics are either a sample of the upper mantle or provide a useful analogy for what that material is like. Our model may conveniently begin where crystallisation has reached a level, perhaps 200 km from the surface, at which the crystal phases are fairly similar to those with which we are familiar at ordinary pressures. The silicate liquid at this stage is probably still close to forsteritic olivine in composition and we may expect, by analogy with the alpine ultrabasics, that olivine, enstatite[1] and chrome spinel would be the primary precipitate minerals forming over a long period. At a later stage plagioclase and augite would be expected to join the other primary phases and the rock type formed would gradually change from ultrabasic to eucritic and finally to gabbroic, with iron enrichment of the ferromagnesian minerals and soda enrichment of the plagioclase. When the level of the crystal accumulate is about 10 km from the surface, a residual liquid of granodioritic composition would be expected to form, if again we use the analogy of layered intrusions guided by the phase chemistry of the relevant silicate systems.

The development of the mantle as a bottom accumulate leads to the conception of a regular upward change in composition. At present one of the intriguing, unsolved problems is concerned with the homogeneity or otherwise of the mantle. Convection of the solid mantle, to which are sometimes attributed the major features of the earth's surface, cannot occur if compositional changes cause differences in density exceeding very small amounts. Inversions to higher pressure crystal phases, if not too sluggish, would not interfere with the convective process. At the later stages of any fractionation process changes in composition are rapid compared with the earlier. In the postulated later layered rocks forming the outer part of the earth the compositional changes might well be considerable and such as to prevent

[1] This is assuming that at these pressures olivine and enstatite have a cotectic, and not a reaction relationship. If this proves to be correct then it will considerably affect the sequence of layered rock types produced. At the level at which the cotectic gives place to the reaction relationship, a harzburgite will give place to dunite which will continue to form until a considerable thickness has accumulated and it should then be replaced abruptly by some olivine-free basic rock which may have become part of the crust.

happened in less than a million years (Jeffreys, 1952, p. 269), although possible factors which could not be evaluated might have considerably extended the time.

Any crystal accumulate, at the time of its formation, is surrounded by liquid. From simple experiments with lead shot and other pieces of metal of various shapes immersed in water, the amount of liquid relative to crystals may be seen to lie between 40 and 60 per cent of the whole. In the case of the layered intrusions so far studied the behaviour of the interstitial liquid during further cooling is variable and far from obvious; in the development of the earth's mantle it is likely to be even more complicated and at the same time of especial significance.

In the Skaergaard layered intrusion the bulk of the liquid trapped between the crystals of the primary precipitate, crystallised out slowly over a considerable range of temperature. There was little reaction between the liquid and the early formed crystals to establish equilibrium between them at the lower temperatures; instead the interprecipitate liquid mainly formed lower-temperature solid solution zones round the original crystals. In the Rhum layered intrusion described by Brown (1956) and perhaps also in the Stillwater layered intrusion being described by Hess (in the press), the primary precipitate crystals grew by diffusion of the required substance from the overlying liquid while they lay within a few inches of the temporary top of the crystal pile. The extent to which the diffusion effect operates depends on the rate of accumulation of the crystal precipitate and on other factors such as the ease of nucleation of the crystals. Growth of the primary precipitate crystals as they lie at the bottom of the liquid requires removal of heat. In the silicate shell of the earth, as in large intrusions, the heat loss must have been through the overlying liquid. For this to take place without producing a new shower of crystals from the liquid, I suggest that there must have been delay in crystal nucleus formation resulting in some degree of supersaturation. At the time of the crystal accumulation to form the mantle, various processes may have reduced the amount of the pore liquid but it is unlikely that they would succeed in eliminating it altogether. The original mantle of the earth should consist, therefore, of the crystal phases appropriate to the temperatures and pressures obtaining together with a certain amount of the contemporaneous liquid.

The composition of any pore liquid trapped between the primary precipitate crystals will have the composition of the liquid phase existing at that time and will be different from that of the contemporary crystal precipitate. Crystallisation of the pore liquid will take place over a range of temperature as cooling slowly proceeds; in the case of the mantle the time required may be measured in millions of years and indeed it may not yet be complete because of the high temperature still obtaining. In the case of the accumulate forming the mantle there would probably be time enough for the original precipitate to be made over by reaction with the liquid to solid solutions in equilibrium at the lower temperatures.

While the crystal phases separating as an accumulate will contain only such chemical elements as can enter the particular crystal structures, the trapped interprecipitate liquid contains all the elements of the contemporary liquid phase. Among these will be the bulk of the naturally radioactive elements such as uranium, thorium and potassium and thus the amount of pore liquid becomes a vital factor in the heat balance of the crust.

By analogy with layered intrusions we may expect to understand the behaviour of a bottom accumulation of crystals up to a thickness of a few kilometres. If, however, the pile of loose crystals, surrounded by a liquid of lower specific gravity, is a hundred or so kilometres thick, the excess of weight of the column of crystals over the column of liquid will become an important factor in the situation. In the deeper part of the column the crystals would be expected to go into solution where in contact with each other and to reprecipitate elsewhere, so that the rock should develop a metamorphic texture. Any disintegration of the crystals, due to higher

pressures in the crystal network than in the surrounding liquid, should produce a tendency for the upward migration of the liquid relative to the precipitate. These two processes only emerge as significant in thick crystal accumulates and they have not been certainly shown to occur in the known layered intrusions. They should, however, be operative and highly important in the mantle, except perhaps in the uppermost part. The extent to which the original interprecipitate liquid of the mantle has been eliminated may be decided, one day, from the heat balance of the earth because the heat producing radioactive elements must originally have been present mainly in the interprecipitate liquid. The postulated interprecipitate part of the mantle is just what is required in Wilson's hypothesis (1957) on the production of the earth's crust by the expulsion from the mantle of the part having lower temperatures of crystallisation, a process which he envisages as taking place continuously throughout the whole of geological time.

The question of the actual crystal phases separating in the mantle as a bottom accumulate cannot be decided by direct observation at present, nor is it within the scope of present experimental techniques. We are, however, on the threshold of important advances, for pressures corresponding to a depth of 250 km within the earth are being attained in the laboratory under fairly well controlled conditions. All that can be said at present is that the crystal phases separating in the lower mantle, including perhaps a cubic modification of olivine, as first suggested by Bernal (1936), are such as to lead to a liquid capable of giving rise to the upper mantle and crust.

In speculating on the later stages of fractionation of the silicate liquid to give the upper part of the mantle and crust, we shall make use of the analogy with layered intrusions and shall also assume that the alpine type of ultrabasics are either a sample of the upper mantle or provide a useful analogy for what that material is like. Our model may conveniently begin where crystallisation has reached a level, perhaps 200 km from the surface, at which the crystal phases are fairly similar to those with which we are familiar at ordinary pressures. The silicate liquid at this stage is probably still close to forsteritic olivine in composition and we may expect, by analogy with the alpine ultrabasics, that olivine, enstatite[1] and chrome spinel would be the primary precipitate minerals forming over a long period. At a later stage plagioclase and augite would be expected to join the other primary phases and the rock type formed would gradually change from ultrabasic to eucritic and finally to gabbroic, with iron enrichment of the ferromagnesian minerals and soda enrichment of the plagioclase. When the level of the crystal accumulate is about 10 km from the surface, a residual liquid of granodioritic composition would be expected to form, if again we use the analogy of layered intrusions guided by the phase chemistry of the relevant silicate systems.

The development of the mantle as a bottom accumulate leads to the conception of a regular upward change in composition. At present one of the intriguing, unsolved problems is concerned with the homogeneity or otherwise of the mantle. Convection of the solid mantle, to which are sometimes attributed the major features of the earth's surface, cannot occur if compositional changes cause differences in density exceeding very small amounts. Inversions to higher pressure crystal phases, if not too sluggish, would not interfere with the convective process. At the later stages of any fractionation process changes in composition are rapid compared with the earlier. In the postulated later layered rocks forming the outer part of the earth the compositional changes might well be considerable and such as to prevent

[1] This is assuming that at these pressures olivine and enstatite have a cotectic, and not a reaction relationship. If this proves to be correct then it will considerably affect the sequence of layered rock types produced. At the level at which the cotectic gives place to the reaction relationship, a harzburgite will give place to dunite which will continue to form until a considerable thickness has accumulated and it should then be replaced abruptly by some olivine-free basic rock which may have become part of the crust.

any convection currents in the solid mantle from invading the uppermost layers which are now the crust. Whether the slighter compositional changes produced during the supposed bottom accumulation would allow convective overturn of the solid mantle is an open question.

The hypothesis just envisaged results in an outermost sialic crust over the whole of the earth, yet now the sial is concentrated to form the continents. By the time the residual sialic liquid developed there would be only 10 km or so depth of liquid left and the convection cells in a thin layer of this kind would be only of similar lateral extent. Thus convection at this stage could not be responsible for pushing the sial into its present, restricted area. The problem of how the sial comes to occupy only a third of the earth's surface is a problem which has often been considered but without arriving at any very plausible solution and perhaps I may be allowed to put forward another possibility. As originally accumulated, the mantle would contain a considerable amount of interprecipitate liquid. This would affect the physical properties and the amount of radioactive heating. Perhaps in an intermediate stage, and before the solidification of an overlying low-melting granodioritic liquid, a major convective overturn took place in the mushy mantle, pushing the sialic scum, already formed by the crystal fractionation, into the limited areas it now occupies. It is possible that convective systems have developed in the earth at three different stages of its evolution: first in the liquid stage, then at a stage when the material consisted of crystals with much pore liquid and finally, at the present stage, when the mantle is essentially crystalline.

I must not attempt to follow up other lines of thought except to refer once more to the basalt problem. As I see it, basalt magma, in general, is likely to be derived from partial melting of peridotite below the Mohorovičić discontinuity. There are, however, two major types of basalt magma which, it is pleasant to recall, were first defined from the Scottish Tertiary igneous province. During the mapping of Mull, the Survey geologists distinguished a Plateau Magma type of basalt (since called by Tilley and others *alkali basalt*) and the Non-porphyritic Central type of basalt (since called by Kennedy *tholeiitic basalt*). Although intermediate types of basalt exist, the distinction discovered in Mull is found to be of world-wide significance and the ultimate origin of the two major types has been much considered. Recently Kuno (1957) has broken away from the older approaches to the genetic problem and suggested that both basalt magma types are the result of partial melting of the mantle, but producing different results because the melting took place at different depths. He suggests that the incongruent melting of forsteritic olivine which occurs at low pressures is replaced at higher pressures by normal congruent melting, the olivine and enstatite then having a cotectic relationship. The texture of some alpine type harzburgites, it may be noticed, seems to indicate a cotectic rather than a reaction relationship between olivine and enstatite. Assuming that the upper mantle consists of olivine and enstatite then the product of slight melting at high levels, where the reaction relation holds, will produce a liquid of tholeiitic type, while the product of melting at greater depths, where cotectic relationships prevail, will be of Plateau Magma type. Partial melting of the intermediate layer of the ocean areas, if it be of gabbro composition, should result in andesitic or more acid lavas. Hess (1955, p. 431) has suggested that the andesitic lavas of the West Indies are due to melting of gabbroic material forming a downward bulge of the crust, a structure which is indicated by gravity and seismic observations.

In a solid mantle, particularly if the temperatures increase with depth as slowly as is now generally considered, it is something of a problem to understand how basaltic magma can be produced at all. However, volcanic action is characteristically periodic and therefore we need not usually postulate more than a small amount of basalt magma being in existence at any one time. Even the larger gabbroic intrusions, such as the Bushveld of South Africa, are very small compared with the

whole mantle. Thus it looks as though we may be able to think of basalt as derived by the squashing out of relatively small amounts of liquid from a large amount of mantle. Whatever the mechanism involved in getting liquid basalt from the mantle—and several possibilities can be visualised—it will leave the mantle locally altered in composition; once the low temperature liquid has been pressed out, the mantle will locally become barren, to use Bowen's expression (1928, p. 317). Some variation in the upper mantle besides that due to layering is therefore likely, and has perhaps already been detected seismologically in the Caribbean. Where the mantle has been rendered barren it would be expected to have metamorphic rather than accumulative textures. If, however, some low-temperature pore liquid remains, at any rate in the upper mantle, we may have an explanation of the asthenospheric conditions for which Daly so strongly argued (1948). Any liquid pore material must also affect other physical properties of the mantle and in particular it should decrease the velocity of the earthquake waves, a matter which might be investigated experimentally.

The chief features of the proposed compositional model may perhaps be summarised in this way. Assuming at some early stage a molten earth (and probably heating by the radioactivity of ^{40}K would ultimately ensure this if its composition were that of the chondritic meteorites) then convection seems inevitable; as a result there should be bottom accumulation of crystal phases giving solidification of the mantle from the bottom upwards. Analogy with layered intrusions provides useful general ideas on the manner of accumulation of the crystal phases, and the physical chemist working on silicate systems at high pressures should soon be able to indicate the actual phases crystallising, providing the overall composition is sufficiently well known. However, just as the sinking of crystals in a crucible in the laboratory provides an inadequate picture of what happens during the accumulation of crystals in a layered intrusion, so new features, due to the greater scale, will undoubtedly arise when the pile of loose crystals produced by bottom accumulation is of the order of hundreds instead of tens of kilometres in thickness. By its own weight the pile of accumulated crystals would be expected to disintegrate, giving a sort of metamorphic rock and expelling, during the process, much of the interprecipitate liquid—at any rate from the lower layers. The amount of interprecipitate liquid will also complicate the question of convection within the earth, partly by the heating effect of the radioactive constituents and partly by affecting the mechanical properties. At an early stage in earth history there was probably easy and rapid convection in the molten silicate part, then there may have been an intermediate stage when convection took place in a mantle which had still a considerable proportion of uncrystallised liquid between the primary precipitate grains, and finally there is the possibility of extremely slow convection in the crystalline mantle. Analogy with layered intrusions forces our attention on the existence, at least temporarily, of an interprecipitate liquid. How this has been, or is still being, eliminated is a matter of great interest.

Concluding Remarks

A deep boring into the rocks below the crust will enable direct determinations to be made of the velocities of compressional earthquake waves in the material encountered and it will provide detailed information on the temperature gradient. Such data will give greatly increased confidence in extrapolating from our usual surface measurements. The samples of rock obtained will allow petrologists and geochemists to answer many of the questions raised in this review. At the first attempt the bore hole may well not reach its deep objective, but on the way many discoveries will be made; we may find, for instance, much of the earth's history preserved in the successive layers of sediment on the ocean floor and we should

also find out the nature of the so-called basalt or intermediate layer. Costly and difficult as the drilling down to 10 km will be, we shall soon begin, I expect, to consider the possibility of borings to even greater depths in continental areas. For the successful deep drilling of the earth, pooling of knowledge, and considerable sums of money will be needed. The project is thus an obvious one for international scientific co-operation.

I have chosen a subject for my address which links fields of science conventionally classed as geology, geophysics, and geochemistry. I do not think of them in this way but regard geology as embracing both the physical and chemical approaches to the study of the earth. The content of what is called geology varies with the possibilities existing at the time, particularly developments in the more fundamental sciences—physics, chemistry and biology. The plan to bore a hole to see what lies below the crust of the earth is largely due to those who call themselves geophysicists, but it is following the best geological tradition. Lyell, in his *Principles* (1830, p. 60), quoted with approval Desmarest, who was in the habit of answering geological arguments by the words, 'Go and see'.

The background for this review is the work of many geologists, especially those interested in the physical and chemical aspects. I particularly acknowledge help from the writings of Birch, Bowen, Bullard, Daly, Hess, Jeffreys, Kuno, Ross, Urey and Verhoogen. Three classic works and some of the more specialised sources of information are listed below.

References

Classic Works

Bowen, N. L. (1928): *The evolution of the igneous rocks*. Princeton.

Daly, R. A. (1933): *Igneous rocks and the depths of the earth*. New York.

Jeffreys, H. (1929 (2nd Edition) and 1952 (3rd Edition)): *The earth: its origin, history and physical constitution*. Cambridge.

Specialised Sources

Adams, L. H. (1924): 'Temperatures at moderate depths within the earth', *J. Wash. Acad. Sci.* **14**, 459–72.

Benson, W. N. (1926): 'The tectonic conditions accompanying the intrusion of basic and ultrabasic igneous rocks', *Nat. Acad. Sci. (Washington)*, **19**, 1st Mem., 1–90.

Bernal, J. D. (1936): Geophysical Discussion, *The Observatory*, **59**, 268.

Birch, F. (1954): 'The earth's mantle: elasticity and constitution', *Trans. Amer. geophys. Un.* **35**, 79–85.

Bowen, N. L. & Tuttle, O. F. (1949): 'The system MgO-SiO_2-H_2O', *Bull. geol. Soc. Amer.* **60**, 439–60.

Brown, G. M. (1956): 'The layered ultrabasic rocks of Rhum, Inner Hebrides', *Phil. Trans.*, Ser. B, **240**, 1–53.

Buddington, A. F. (1943): 'Some petrological concepts and the interior of the earth', *Amer. Min.* **28**, 119–40.

Bullard, E. C., Maxwell, A. E. & Revelle, R. (1956): 'Heat flow through the deep sea floor', *Advanc. Geophys.* **3**, 153–81.

Daly, R. A. (1946): 'Nature of the asthenosphere', *Bull. geol. Soc. Amer.* **57**, 707–26.

Davis, G. L., Aldrich, L. T., Tilton, G. R., Wetherill, G. W. & Jeffery, P. M. (1956): 'The age of rocks and minerals'. In the Annual Report of the Director of the Geophysical Lab., Carnegie Inst. of Washington, pp. 167–8.

Du Rietz, T. (1935): 'Peridotites, serpentines, and soapstones of northern Sweden', *Geol. Fören. Stockh. Förh.* **57**, 133–260.

Hess, H. H. (1938): 'A primary peridotite magma', *Amer. J. Sci.* **35**, 321–44.

Hess, H. H. (1955A): 'Serpentines, orogeny, and epeirogeny', *Geol. Soc. Amer. Special Paper 62*, pp. 391–408.

Hess, H. H. (1955B): 'The oceanic crust'. Sears Foundation: *J. Mar. Res.* **14**, 423–39.

Hill, M. N. (1957): 'Recent geophysical exploration of the ocean floor'. *Phys. Chem. Earth*, **2**, 129–63.

Kelvin, Lord (1897): 'On the age of the earth as an abode fitted for life'. Annual Address of the Victoria Institute, London.

Kuhn, W. & Rittmann, A. (1941): 'Über den Zustand des Erdinnern und seine Entstehung aus einem homogenen Urzustand', *Geol. Rdsch.* **32**, 215–56.

Kuno, H., Yamasaki, K., Iida, C. & Nagashima, K. (1957): 'Differentiation of Hawaiian magmas', *Jap. J. Geol. Geogr.* **28**, 179–218.

Kuno, H.: 'Origin of Cenozoic petrographic provinces of Japan and surrounding areas'. In press.

LACROIX, A. (1893): 'Les enclaves des roches volcaniques', *Ann. Acad. Mâcon*, Ser. 2, **10,** 1–710.
LACROIX, A. (1894): 'La Iherzolite des Pyrénées', *Nouv. Arch. Mus. Hist. nat., Paris*, 3e series, VI.
LACROIX, A. (1942): 'Les péridotites de la Nouvelle Calédonie: leurs serpentines et leurs gîtes de nickel et de cobalt', *Mém. Acad. Sci., Paris*, **66,** 1–143.
OLDHAM, R. D. (1906): 'The constitution of the interior of the earth, as revealed by earthquakes', *Quart. J. geol. Soc.* **62,** 456–75.
RINGWOOD, A. E.: *On the chemical evolution and density of the planets*. In press.
ROSS, C. S., FOSTER, M. D. & MYERS, A. T. (1954): 'Origin of dunites and of olivine-rich inclusions in basaltic rocks', *Amer. Min.* **39,** 693–737.
TILLEY, C. E. (1947): 'The dunite-mylonites of St. Paul's Rocks (Atlantic)', *Amer. J. Sci.* **245,** 483–91.
UREY, H. C. & CRAIG, H. (1953): 'The composition of the stone meteorites and the origin of the meteorites', *Geochim. et cosmoch. Acta*, **4,** 36–82.
VERHOOGEN, J. (1956): 'Temperatures within the earth', *Phys. Chem. Earth*, **1,** 17–43.
WAGER, L. R. & DEER, W. A. (1939): 'Geological investigations in East Greenland. Part III. The petrology of the Skaergaard intrusion, Kangerdlugssuaq', *Medd. Grønland*, **105,** No. 4, 1–352.
WILSON, J. T. (1957): 'Origin of the earth's crust', *Nature, Lond.* **179,** 228–30.

SOME BIOLOGICAL EFFECTS OF LOW TEMPERATURES[1]

ADDRESS BY DR. A. S. PARKES, C.B.E., F.R.S.

PRESIDENT OF SECTION D

Introduction

The study of the biological effects of cold has attracted many of those prominent in the history of science. In 1689, before the invention of an accurate thermometer, or of especial methods of producing extreme cold in the laboratory, Robert Boyle recorded a series of most intriguing observations on problems ranging from the preservation of food at low temperatures to the effects of cold on knights in armour. More relevant to the present discussion, he observed that frogs and fish were viable after being encased in ice for a day or two but not after longer periods. Réaumur, in 1736, used his newly invented thermometer to facilitate observations on the effects of cold on a number of different organisms, notably insects. More extensive foundations to the subject were laid in 1787 by Spallanzani who used mixtures of ice and various salts to produce freezing mixtures and who studied the effects of temperatures down to 24° below zero on animalcules, insects, fish, amphibia, reptiles, birds and mammals, and on their ova and spermatozoa. In the course of his work, Spallanzani made the far-reaching observation that supercooling was less damaging than freezing to unicellular organisms.

In the 150 years that followed, the introduction of solidified CO_2 and liquid gases greatly increased the range of temperatures over which experiments could be carried out, and work was extended to a much greater range of plants and animals and their cells and tissues. In 1940, Luyet and Gehenio gave a broad historical review in their book *Life and Death at Low Temperatures*.

Since then, progress has been very rapid, both with the whole animal, especially the mammal, and also with isolated cells and tissues. This progress can be ascribed to the introduction of reduced body temperature as an aid to human surgery, to the current importance of studies on cold injury, and finally to the discovery of the protective action of glycerol and other neutral solutes against the otherwise fatal effects of freezing and thawing on isolated mammalian cells and tissues.

In the face of this wealth of material I must define and limit my subject more closely than can be done by a brief title. I shall deal only with animals, and among animals only with multicellular organisms. In making this limitation I do not imply that knowledge of temperature effects on plants and unicellular organisms is negligible, but that it is too extensive to be included. In considering multicellular animals I shall consider both the whole animal and isolated cells and tissues, but I shall make only passing reference to aspects being dealt with specifically by later speakers in this symposium. A second limitation of my theme is that I use the expression 'low temperature' in relation to the internal temperature of the cell, tissue or organism rather than to environmental temperature, the latter being important only in so far as it affects the former.

Cooling and Freezing of Whole Animals

Cold-blooded animals

According to the traditional view, cold-blooded animals have little if any control over their internal temperature, which tends to approximate to that of the environment. In outline, this traditional view is self-evident. Poikilothermic animals die

[1] Address delivered on August 28, 1958, at the Glasgow Meeting of the British Association.

rapidly when their body temperature rises above about 45°C, but many withstand readily body temperatures, reached in adjustment to environmental temperatures, down to 0°C, at which their vital processes are slowed down but do not cease. In general, the higher poikilothermic animals are characterized by the fact that their vital functions have a higher tolerance for temperature changes, but it is now thought that many of them have some power of regulating their internal temperatures and it may be that such a mechanism is important when the animal is in danger of being frozen.

Experiments have been carried out on a great variety of the lower poikilotherms; Luyet, for instance, succeeded in reviving vinegar eels cooled to −190°C. For the most part, however, study of the poikilotherm at body temperatures below zero has been concentrated on three main groups, insects, molluscs and fish. The insects will be dealt with later in this session by Dr. Mellanby. All that need be said here is that many insects or their larvae are regularly subjected to environmental temperatures far below 0°C, with a corresponding reduction in their internal temperatures. Under experimental conditions, larvae and pupae of certain insects have shown a remarkable capacity for supercooling, even to temperatures as low as −30°C, and this phenomenon undoubtedly plays a large part in their frost hardiness. Nevertheless, there are now numerous reports of insects surviving the crystallisation of a large part or even the whole of their body water. Asahina, in a personal communication, tells us that he has revived moth larvae which had been frozen to −30°C and then stored in liquid air.

Sea-water freezes at about −1·7°C and this is therefore the lowest environmental temperature to which most marine poikilotherms are normally subjected. Intertidal organisms, however, may encounter much lower temperatures, and investigations have shown that various molluscs have a noteworthy toleration for the crystallisation of their body water as ice, and for the hypertonicity of their body fluids resulting therefrom.

Probably the most widespread of all the legends of cryobiology is that fish have been revived after being frozen stiff in blocks of ice. This idea has arisen partly from the fact that fish will survive for a time freezing of the surface of the body, and partly from the fallacious assumption that a fish encased in ice is itself necessarily frozen. It is certainly not true in the sense that fish will survive total freezing of the whole body under any conditions so far studied. Nevertheless, many illuminating experiments have been carried out on fish at sub-zero temperatures and we are fortunate in having Dr. Kanwisher with us to deal with this aspect of our subject.

Warm-blooded animals

In contrast to the cold-blooded animals, the warm-blooded ones typically regulate their internal temperatures within close limits. This category includes only mammals and birds, and zoologically, therefore, may be considered of minor importance. Nevertheless, they are physiologically of great interest because of the delicacy of the regulatory mechanisms and the complexity of the supplementary devices. The regulated temperature varies somewhat from species to species, but is almost always between 37°C and 41°C. In most parts of the world therefore, it is much above environmental temperature, and in some instances the gradient may be astonishing. The Emperor Penguin, with a body temperature about 40°C, supports life in antarctic temperatures approaching −60°C and, moreover, contrives in such conditions to incubate eggs and care for the chicks. Arctic mammals have similar problems. In a less spectacular way, one may wonder how a house sparrow, weighing only an ounce or two manages to support a temperature difference of 40°C between itself and its winter environment.

Devices for conserving body heat are, of course, numerous. Insulation is achieved by layers of feathers, fur or fat, and heat loss may be minimised by local poikilothermy of the extremities. Moreover, many young animals and birds, which are especially liable to exposure, show a marked capacity for poikilothermy. Another device, used by many adult mammals and apparently some birds, is that of hibernation during periods of cold or food shortage. In this condition the body temperature falls to, and is controlled at, a much lower level as though a thermostat had been re-set. At the same time vital activity is reduced. The cycle of hibernation is usually seasonal, but in extreme cases may be diurnal. Certain bats and humming birds, for instance, are said to hibernate each night. Seasonal hibernation in an adult homeothermic animal is normally preceded by a period of preparation. The animals put on weight or store food. The body fats become unsaturated so that their melting point is reduced, and physiological changes take place which enable the nervous and other systems to function at a lower temperature. It should be emphasised that the hibernating animal is not in a state of suspended animation. Heart-beat and respiration continue at a slow rate. Moreover, the capacity for thermal regulation is not lost entirely, and unless it is moribund or under experimental conditions the hibernating animal does not allow its body temperature to fall to a level at which body water is converted into ice. Normally, the body temperature remains at a few degrees above 0°C, but in the extreme case of certain bats, as described by Kalabukhov, a remarkable capacity for supercooling enables the heart to continue beating and viability to be retained at a body temperature several degrees below zero. There is no evidence that normal hibernation in mammals and birds ever involves complete arrest of heart-beat or respiration or that under natural conditions they ever recover from the frozen as opposed to the supercooled state.

As regards laboratory experimentation, it was commonly thought until a few years ago that an adult non-hibernating mammal could not survive being cooled to a deep body temperature below 15°C, at which point heart and respiration fail. This view was exploded by the work of Andjus, and it is now known that rats, mice and dogs can be revived after cooling to 0°C, although heart-beat and respiration have been arrested for an hour or more. Even more remarkable is the fact that rats so treated show no impairment of memory or learning capacity. Moreover, Audrey Smith has found that the golden hamster, probably a hibernator in nature, can be resuscitated undamaged after supercooling to −5°C or freezing to the extent of crystallising 50 per cent of its body water. Dr. Smith will herself elaborate this theme during the symposium, but one further point may be mentioned here.

At 0°C the artificially cooled mammal has every appearance of being dead, and by all clinical criteria is dead. Yet, under appropriate conditions it can be revived. This situation poses the intriguing question of how to define death. To the biologist only one answer is possible; death is the state from which resuscitation of the body as a whole is impossible by any currently known means. The two qualifications 'the body as a whole' and 'known means' are important. Individual tissues may remain viable, as shown by their capacity to form a functional graft in another animal, long after the clinical death of the organism, and as knowledge increases the means of resuscitation will extend, so that what constitutes death today will not necessarily constitute death tomorrow. But at present a much lowered body temperature is an essential feature of the kind of suspended animation envisaged above. At normal temperature the body as a whole cannot survive arrest of heart-beat and respiration for any great length of time, though resuscitation after short periods of inanimation on the operating table and in other circumstances has been described. The possibility exists, however, that victims of cold and exposure, associated with a comparatively rapid cooling of the body, may sometimes be regarded prematurely as dead. Whether resuscitation in such cases, if it could be

effected, would be justifiable in view of the possibility of permanent damage is debatable, but the work on experimental animals suggests that this hazard is not as great as might have been expected.

Freezing of Isolated Cells and Tissues

Much of the work on freezing isolated cells and tissues was carried out on material from Invertebrates, and, as already indicated, considerable success was achieved. Normal material from Vertebrates proved to be much more sensitive to the effect of freezing and thawing, and up to 1949 there were only very few reports of even partial revival of cells and tissues of the higher animals after freezing. In that year Smith and Polge observed that the addition of glycerol to the suspending medium had a remarkable effect in protecting fowl spermatozoa against the otherwise fatal effects of freezing and thawing. This effect can be demonstrated very easily. A suspension of fowl spermatozoa is divided into two parts, each of, say, 1 ml.; one part is further diluted with an equal amount of Ringer solution, the other part is diluted with an equal amount of Ringer solution containing 20 per cent glycerol, so that the final concentration of glycerol is 10 per cent. Both tubes are then plunged into alcohol cooled to −79°C by the addition of solid CO_2, and left frozen at this temperature overnight. On thawing next day, it will be found that the spermatozoa in the medium containing glycerol show vigorous motility while those in the other tube are all dead. It should be noted that under such conditions freezing is only moderately rapid, averaging less than 1°C per sec.

Exactly the same kind of protective action of glycerol was found by Smith in the case of rabbit and human red blood cells, which normally are haemolysed immediately by freezing and thawing. Observation under the microscope showed that on addition of the glycerol-containing medium the red cells first shrank a little as water was drawn out more quickly than the glycerol was taken in, and then returned to normal appearance as equilibrium was established. On returning to a normal medium the glycerolised cells swelled as water was drawn in more quickly than the glycerol escaped, and finally burst. It was necessary, therefore, to deglycerolise the cell slowly by dialysis or other means before the frozen and thawed cell could be returned to the body. This observation suggested that the apparent loss of fertilising power by glycerolised fowl spermatozoa, frozen or unfrozen, was due to their being destroyed in the normal fluids of the oviduct, and that the glycerol should be removed slowly before insemination. Subject to this precaution, fowl spermatozoa and human red blood cells, as shown by Polge and by Sloviter and Mollison, behave normally in the body after freezing and thawing in media containing glycerol.

Soon after the discovery of the protective properties of glycerol, attention was turned to the freezing of bull semen, partly because of the abundance of material available and partly because of its economic importance. It was soon found that, even in the presence of glycerol, bull spermatozoa would not tolerate being frozen by the method effective with fowl spermatozoa and red blood cells. After much experimentation a far slower rate of cooling in the early stages was employed, so that the temperature dropped at the rate of only 1° to 2°C per min down to −15°C and thence more rapidly to −79°C. With the technique modified by the introduction of slow freezing, a procedure contrary to all the then existing ideas, a high proportion of bull spermatozoa survived freezing and thawing in glycerol-containing media and were found by Polge and Rowson to have normal fertilising capacity even if the glycerol was not removed before insemination. Recent developments in the technique of freezing spermatozoa will be described by Dr. Polge.

From the early experiments with glycerol, two principles emerged; the use of an effective protective substance, which might or might not have to be removed after thawing before the cells could be used, and the desirability of slow cooling. This

technique, or modifications thereof, has now been applied to many other cells and tissues. Among isolated cells, the mammalian ovum *in vitro* has so far proved very sensitive both to the presence of glycerol and to freezing. The ovarian oöcyte is more resistant, but even so, is less resistant than the endocrine tissue of the ovary, which by grafting experiments in rats and mice can be shown to be functionally viable after slow freezing in the presence of glycerol to −79°C or −190°C. Other mammalian tissues which have been shown to be protected to a greater or lesser extent by glycerol from the otherwise fatal effects of freezing and thawing, include: testis—interstitial cells and seminiferous tubules, adrenal cortex, thyroid gland, cornea, bone marrow, and various accessory reproductive organs, all tested by grafting experiments. Later in this symposium Dr. Barnes will describe experiments with bone marrow. Ovarian granulosa cells, adenohypophysial tissue, renal tissue and cartilage have all been found viable in tissue culture after freezing to −79°C in glycerol-containing media. Nerves, including their cells, fibres and ganglia, transmit impulses normally after similar treatment.

Nature of the Damage caused by Freezing

A rational approach to the problem of avoiding the damage caused by freezing depends on adequate knowledge of the nature of that damage. So far our knowledge is very inadequate. It is an interesting side-light on biological research that so far progress in this field has mainly taken the form of empirical discoveries which later have to be explained.

The simplest concept of freezing damage is that it arises from mechanical pressures and distortions caused by the formation of ice crystals. This concept implies that to avoid freezing damage it would be necessary to avoid crystal formation and it led Luyet to develop methods of ultra-rapid freezing which would vitrify rather than crystallise the water. These methods, applicable only to minute amounts of material, had a limited success, but probably for other reasons than avoiding mechanical damage. It is now known, for instance, from the work of Meryman and of Deanesly, that tissue which is found to be viable after thawing may have been grossly distorted in the frozen state.

A far more serious hazard to frozen tissues was emphasised by Moran. As water crystallises out as ice, the residual fluid becomes more concentrated, and the tissue is literally 'pickled in its own brine'. In more scientific terms, the lipoproteins of the membrane, and probably other components of the cell are destroyed by the hypertonic residual fluids. No doubt other factors are involved. It is known, for instance, that some cells are very sensitive to thermal shock caused by a sudden drop of temperature even in the range above zero and the same phenomenon may occur below zero. However, following the work of Lovelock, hypertonicity of the residual fluid is probably the most important single cause of freezing damage. The effect of 'vitrification', such as it is, is probably due to the avoidance of hypertonic residual fluid. Freezing at the rate of 100°C per sec can be achieved only with very small amounts of, or with highly dispersed material, so that its usefulness is limited. By contrast slow freezing advocated originally by Breedis probably has opposing effects. The cells remain longer at the dangerous temperature where they are exposed to hypertonic residual fluid, but the progressive dehydration of the cell as the extracellular water freezes will decrease the probability of intracellular crystallisation, which according to some authorities is fatal to the cell. It must be supposed that the effect of glycerol is to make the hypertonic residual fluid less damaging to the cell without affecting whatever beneficial effects slow cooling may have on the location of the ice crystals and the avoidance of thermal shock.

In the light of recent work on the nature of freezing damage and on the use of protective substances, insects are of especial interest. The larvae of the goat moth, even in England, often have to withstand temperatures well below zero, and it seems

likely that their remarkable resistance to freezing is due to the substitution of sugars and amino acids for salts and desaturation of the fats during the winter. Even more dramatically, following the laboratory discovery of the protective effect of glycerol, insects of several different species in U.S.A., Canada and Japan have been found to contain glycerol.

Freeze-drying of Viable Material

At temperatures low enough to cause the crystallisation of all intracellular and extracellular water, the material is, in effect, not only frozen but dry. It should therefore make no difference to the cell whether the crystals are left in position or removed. To the experimenter, however, the difference is very real, because removal of the crystallised water makes it possible to keep the freeze-dried material at normal temperatures. This principle has been most successfully applied to many micro-organisms, which, after freezing and vacuum sublimation of the ice, can now be kept for long periods at normal temperatures in a viable condition. This technique, however, presupposes that the cells will withstand not only freezing and sublimation of the water from the crystalline state, but also subsequent reconstitution by the replacement of water. The latter is probably the main difficulty. So far no success has been obtained with mammalian tissues; the freeze-dried artery and bone preparations used as implants in surgical work are, of course non-viable.

In the case of cells and tissues which on freezing are stable only at very low temperatures and with the aid of a protective substance such as glycerol, further difficulties arise. Sublimation of ice, becomes progressively more difficult at decreasing temperatures and is impracticably slow at −79°C. At higher temperatures, e.g. −40°C, sublimation is more rapid, but the time for which the material remains viable is shorter. Further, the removal of the ice crystals will leave residual glycerol in a concentration so high as to be toxic at normal temperatures. For freeze-drying, the protective substance must be volatile at low temperatures, or non-toxic in high concentration at normal temperatures, as well as giving effective protection over a wide range of temperatures. No such substance is at present known, but there is little doubt that a solution to this problem will be forthcoming. Claims have already been made by Russian workers to have freeze-dried bull sperm in such a way as to retain viability.

Long-term Preservation of Cells, Tissues and Animals

Biologists have many reasons to be interested in the effects of cold, but much of the work in this field has been inspired by the possibility of long-term preservation of living material. At very low temperatures, the biochemical processes of life are presumably arrested or greatly retarded, so that if a cell can be frozen and thawed without damage, at the low temperature it should be in a state of suspended animation. Instances of such a state occurring regularly in nature have already been given, and legend, fairy stories and science fiction have contributed liberally to the literature of anabiosis induced by cold.

Laboratory work of the last decade has in fact opened up exciting possibilities of the long-term preservation of cells *in vitro*, such as spermatozoa for insemination, ova for transference to another animal, red blood cells for transfusion, and tissues and whole organs for transplantation. Stretching away beyond these more or less realistic ideas are thoughts of whole animals, possibly even man himself, preserved indefinitely in a state of animation suspended by cold. Some of these pipe-dreams have already been realised. Speaking at the Edinburgh meeting of the British Association in 1951, I called attention to the laboratory work of Smith and Polge on freezing bull spermatozoa and predicted that the use of frozen semen would

become a routine technique in cattle insemination centres all over the world. It has. Moreover, material 'laid down' following the first successful field experiment in Cambridge more than 6 years ago still retains fertilising power. No doubt the technique will be extended to other animals. There are reports from America of human spermatozoa having retained fertilising power after preservation at −79°C for some weeks. Little imagination is required to see where this development could lead.

Among other isolated cells, human red blood cells have been maintained in condition suitable for transfusion for more than 2 years by means of the glycerol-freezing technique. The same applies to rat ovarian tissue which has been found to form endocrinologically active grafts after 2 years or more preservation at −190° C. Human cornea has been grafted successfully after some months storage at − 79°C. The record of long-term storage of mammalian cells and tissues in viable condition grows each year, but the full exploitation of the technique will take a long time. The study of immortality is necessarily a long-term project.

As regards the whole mammal, or even the whole vertebrate, work on suspended animation is at a very early stage. The possible duration of suspended animation in mammals at ordinary body temperature is severely limited by the rapid breakdown of the elaborate biochemical systems of the body and the onset of irreversible changes. The same is true to a lesser extent even at a body temperature of 0°C. The process of biochemical disintegration following arrest of heart-beat and respiration is slowed down but is not stopped by cooling to this extent. For this and other reasons there is no doubt that the condition of the cooled or partly frozen animal is very unstable and no prolonged survival can be expected at this temperature. For long-term preservation in the frozen state much lower temperatures will be necessary, presumably in the range already known to be required for individual cells and tissues. The induction of suspended animation at −80°C will no doubt necessitate the pretreatment of the mammal with some such substance as glycerol at a concentration which according to present knowledge is highly toxic to the whole animal. In other words, the biologist is not obviously near to achieving suspended animation of a warm-blooded animal at a temperature likely to result in a stable state. But he may do so unexpectedly at any moment by some ingenious stroke.

Conclusion

To the biologist the potentialities of this work on the application of low temperatures to cells, tissues and whole animals are limitless, and developments in this field are likely to constitute a major contribution to man's powers over nature. This prospect is of some historical interest. For centuries biological science has lagged far behind the physical and chemical ones in the understanding and modification of natural processes. What Isaac Newton did for physical science was not done for biological science until two centuries later by Charles Darwin. This lag has been due to many causes. For one thing, religious dogma and private prejudice retained their grip on biological science long after astronomical and physical sciences had escaped; for another, it can be argued that biology, especially experimental biology, is the more difficult discipline, demanding wider knowledge, longer study and more complete differentiation of the mind from the emotions, and is likely, therefore, to come to full flower later in man's evolution. With the liberation of atomic energy, physical science has taken another tremendous stride. There is evidence on every hand, however, that the tempo of biological discovery is quickening dramatically and that man's control over biological processes will not continue to lag behind his control of the physical world. Among this evidence recent work on the biological effects of low temperatures is not the least.

References

The following reviews may be consulted for details and references to the literature:

Lovelock, J. E. (1954): 'Biophysical aspects of the freezing of living cells'. Ciba Foundation Symposium: 'Preservation and Transplantation of Normal Tissues', p. 131.

Luyet, B. J. & Gehenio, P. M. (1940): 'Life and Death at Low Temperatures', *Biodynamica*, Normandy, Miss.

Lyman, C. P. & Chatfield, P. O. (1955): 'Physiology of hibernation in mammals', *Physiol. Rev.* **35,** 403.

Meryman, H. T. (1956): 'Mechanics of freezing in living cells and tissues', *Science*, **124,** 515.

Parkes, A. S. (1956): 'Preservation of living cells and tissues at low temperatures', *Third Internat. Congr. Anim. Reprod.*, Cambridge. June 1956. Plenary Papers, p. 69.

Parkes, A. S. & other authors (1957): 'A discussion on viability of mammalian cells and tissues after freezing', *Proc. roy. Soc.*, B, **147,** 520.

Rey, Louis R. (1957): 'Aspects physico-chimiques de la congélation des tissus et de leur conservation par les basses températures', *Ann. Nutr., Paris*, **11,** 103.

Smith, A. U. (1954): 'Effects of low temperature on living cells and tissues', in *Biological Applications of Freezing and Drying*. Ed. R. J. C. Harris. Acad. Press Inc.

Smith, A. U. (1958): 'The resistance of animals to cooling and freezing', *Biol. Rev.*, **33,** 197.

THE ATLANTIC ENDS OF EUROPE[1]

ADDRESS BY PROFESSOR E. ESTYN EVANS

PRESIDENT OF SECTION E

'THE end of the world is very near': with these words Professor D'Arcy Thompson, father of the revered figure whom many of us remember at meetings of the British Association, began a public lecture delivered in Galway on March 22, 1866. He went on to reassure a startled audience that he was speaking 'geographically, not chronologically', for his subject was—Connemara. Nowadays we dare not so neatly oppose space and time, since we have been taught to regard them as related aspects of reality, but on this occasion I too am expected to speak geographically, and Connemara lies near the centre of my subject. Nor is D'Arcy Thompson's text out of place in this capital city of the Highlands, which bears a Celtic name and whose patron saint is St. Kentigern.

The ends of the Old World are indeed very near, but the other world at which D'Arcy Thompson hinted is also never far away. In these lands on the edge of the Atlantic the supernatural world is part of man's environment and may carry as much conviction as the visible universe. Spiritual values and religious fervour are combined with idealisation of the past, and the idea of progress has had little appeal. The Golden Age lies in the misty past: there is a feeling that the best has been, and that man can do no better than strive to restore what has been lost. Saints are held in more esteem than scientists, and because legend usurps historic record, studies of the cultural landscape are more than usually difficult. In this environment even the earth-bound geologist is tempted to lift up his eyes and turn his fascinated gaze on faery lands forlorn. 'To all of us it may be given', wrote Grenville Cole in his *Topography and Geology of Ireland*, 'to catch far off, against the gold of sunset, some glimpses of the fortunate isles.' We cannot doubt that for many seekers such glimpses have been a source of inspiration, but the scientist must not be led astray by siren voices singing sweetly in a Celtic tongue.

It is partly for this reason that I have avoided the term Celtic in my title. Not only is the word heavy with romantic overtones, but it has been applied variously to a group of languages, a culture, an art-style, a fringe, a twilight, a 'race' and a pony. It is used to designate an early field-system and, in this city, a famous football field. Moreover, the Celtic heritage is but one of many cultural strata which have been deposited in the far west; and while I shall be mainly concerned with some interactions between people and environment in the western coastlands of the British Isles, I also wish to consider these lands as part of an Atlantic zone extending from north Portugal to Norway.

Physical and political fragmentation should not blind us to the affinities of environment and society shared by these tattered ends of Europe. The length of the Atlantic fringe, measured along the great circle between the 40th parallel and the Arctic Circle, is some 2,000 miles, or about the same length as the Mediterranean axis from Gibraltar to Istanbul. On both Mediterranean and Atlantic fronts maritime enterprise has been moving men and their goods and ideas about since prehistoric times. Along the Atlantic coasts consciousness of a common cultural heritage was strongest in the early centuries of Christianity, and Irish legendary history looks back to a time when 'the world' extended 'from Brittany to Norway and from the Orkneys to Spain'. Behind this there lies the still unresolved mystery of movement and trade between the western Mediterranean and Denmark

[1] Address delivered on August 29, 1958, at the Glasgow Meeting of the British Association.

in megalithic times. And in later centuries the pilgrimage to Santiago and the wine trade perpetuated traffic along the fringe.

Today it is at the level of folk life and in the rural landscape that cultural parallels along the Atlantic coasts can most readily be discerned. From Galicia to Galway, and on to the Faeroe and Lofoten Islands the inhabited coastlands, bare of woods but green with grass, are parcelled out in small fields protected by stone dykes, sod walls or hedged banks. Settlements are dispersed, consisting of single farms or small vestigial clusters of farms instead of village communities, and the arable land is interrupted by extensive areas of bare rock and bog, of mountain grasses, heather, bracken and gorse. Urban centres are typically small and of no great antiquity. Economically and culturally the business of rearing livestock is of paramount importance. As compared with the ordered arable farming of village communities a 'familiar' pastoral way of life, less dependent on material equipment and organisation, has bred suspicion of external authority. The blood brotherhood of a clan system—for example—implies a rough equality, and the custom of partible inheritance and communal rights have been jealously guarded. Transhumance has been a means both of utilising hill grazing to provide a healthy variety for livestock, and of conserving fodder and safeguarding crops on the small home farms. It has served a useful purpose in allowing the young people of isolated glens—always in danger of excessive in-breeding—to mingle and intermarry. In times of stress an extended transhumance has facilitated survival among the hills, and the sea has provided an alternative means of movement, of sustenance and escape, for saints and sinners alike. Well may the Welsh, and the Irish (if with less harmony) sing of their hills, and the Hebrideans of their seas. We recall the two mighty voices of Wordsworth, and his idyllic description of the Lake Country—one of our Atlantic Ends—as 'a perfect republic of shepherds and agriculturalists'.

Land-divisions—e.g. townships and townlands—are small, and tools and techniques simple. The traditional dwelling, in many parts of the fringe, is the single-storied long house, only one room wide, built of stones, sods or mud, normally whitewashed and roofed with thatch. Until distributing centres such as Glasgow pioneered in the mass production of oven-bread, the Atlantic Highlanders and Islanders made their own thin-bread, without yeast, from oats, barley or rye. They have their traditional styles of costume and of furniture, such as the mantle and the shawl, the settle and the dresser. Most of the Atlantic Ends share, or have but lately lost, many other culture elements: sledges, two-wheeled carts and swing ploughs, mattocks and narrow spades, long-handled shovels, grain-drying kilns and mortars, rawhide shoes and clogs, alcove beds, open oil lamps, a rich oral literature and a wealth of folk songs, typically using an old pentatonic scale. Family and social life is marked by the strength of the bond of kinship and by the survival of patronymics which are often a source of great pride. These loyalties may override others and deflect the paths of justice and good government.

If one examines these traits it will be found that while some of them have spread by sea from north or south, others were in former times widely distributed in lowland England and on the European mainland: they have found their last refuge in the Atlantic Ends. The first home of a recognisable Celtic language was in central Europe east of the Rhine, whence Celtic tongues spread over western Europe in the last millennium B.C. Their subsequent fate provides a good illustration of the functions of the Atlantic fringe as both a refuge and a path of coastwise diffusion: reaching the British Isles in pre-Roman times, Celtic languages were later carried from Ireland to Scotland and from Cornwall were reintroduced to Brittany. Their survival in our islands, in the face of an aggressive tongue which has conquered continents overseas, is eloquent testimony of the personality of what Sir Cyril Fox has called the Highland Zone of Britain.

It is not my purpose to attempt a summary of the physique of the Atlantic Ends,

but rather to call to mind a repeating pattern of highlands and islands between the Lofoten Islands and the mouth of the Douro. Ancient massifs varying in height from a few hundred to several thousand feet, composed mainly of igneous and metamorphic rocks of great variety, face the Atlantic in deeply indented coasts. The proportion of upland to lowland increases as one goes north, where habitability is further reduced by the lower upward limit of cultivation, but the combination of abundant rains and ancient rocks characterises all the Atlantic Ends, Hercynian as well as Caledonian. Resistant quartzites, granites and gneisses everywhere yield little soil, poor in plant nutrients. On the other hand, even in the north, sheltered pockets where glacial and post-glacial materials have accumulated are of surprising fertility if they contain lime, and the raised beaches lining the sea-inlets of Scotland and Norway are of special significance for man, and provided him with his first safe foothold.

The heavy precipitation, evenly distributed through the year, is associated with mild winter temperatures and high values of wind and cloud. Here I will refer only to the wind as a vital environmental factor, affecting man's affairs in innumerable ways. Farmer and fisherman alike pay it the respect which is owed to the unseen, and turn it to good account. When I commiserated with a west of Ireland farmer on a harvest day of severe gales he replied: ' 'Tis wet, but thank God 'tis windy'. For it is mainly to the wind that the crofter looks to dry his crops, his peats, his salt-fish. The wind brings ashore the drifting wrack which is harvested along all the Atlantic coasts. It blows inland the shell-sands of the machair and has thus enriched islands of low relief which offer no barrier to it, such as Tiree—once the granary of the Isles. Recent research suggests that the Atlantic winds may also contribute plant nutrients derived from atmospheric salts deposited in rain. Some such climatic factor perhaps explains the Clareman's proud boast—allowance being made for regional rivalries—that he can grow three blades of grass to every two in the fat pastures of County Meath. High winds, by discouraging virus-carrying aphids, assist the farmers of Northern Ireland and south-western Scotland in producing healthy stocks of seed potatoes for export. Protection from wind has been a primary factor in the siting and construction of rural dwellings: they tend to face not into the sun but away from the wind. Roofs are streamlined to deflect it, walls and thatch thick to deaden its roar. In single farms and clustered clachans alike buildings are placed to protect one another from the wind. Nor should the social consequences of wild weather be forgotten. In this environment the rigid adherence of the lowlander to a time-table in a programme of work—whether it be a commercial undertaking or a geographical field-trip—is no virtue: it may bring disaster.

The forms of human settlement in any region cannot be explained without reference to the cultural environment in which they took shape, and in our Atlantic Ends, where continuity of peoples and traditions is strong, the geographer is faced with historical questions to which history has been unable to provide answers. For the peasant societies which have constituted the great mass of the population in these lands down to recent times have left few written records, and the ancient laws of their leaders took little notice of humble ways which were rooted in custom. Historians and geographers alike have regarded enclosure and the single-farm system as immemorial features of the Atlantic fringe, and it has been assumed that such traces of open-fields as have been found there were the result of manorial influence. The Scottish infield-outfield scheme is well documented, but H. L. Gray's identification of similar forms of land use in his *English Field Systems* (1915), has not won adequate recognition, and the survival in varying degrees of related patterns of field systems and settlements in all the Atlantic Ends of Europe has escaped notice.

History cannot come to grips with such timeless communities, and the historical geographer, in so far as he is concerned with the rural landscapes of the past, is

without his usual tools and can tell us little of the underlying continuities of crafts and oral traditions, of ritual and festival. For this information we must turn to the folklorist and the prehistorian, for essential elements of the cultural landscape were established with the advent of farming in Neolithic times. And here I must pay a tribute to our archaeological colleagues for the skill with which they have practised the techniques of field research and its expression in distribution maps. Few geographical works are illustrated with the cartographical clarity which is displayed, for instance, in Professor Grahame Clark's ecological studies of prehistoric Europe.

The geographer can also learn much from the geomorphologist. Whatever opinion we may hold as to the precise place of geomorphology in geographical studies, its practitioners have excelled in field research and in mapping techniques. They have long recognised that the interpretation of existing landforms must be sought in the historical processes of earth sculpture, and they seek explanations of present landscapes in the events of 50 million years ago. It is in no spirit of rivalry, but from conviction, that I ask for a span of some 5,000 years and enlist the assistance of fieldwork in an attempt to trace the evolution of a cultural landscape. In Ireland I have been led to study ethnographic survivals and oral traditions in the field, and to dig into the prehistoric past, and I believe that without knowledge thus gained I would have failed to comprehend, and perhaps would not have observed, significant features of present-day life and landscape.

I began my field studies of rural settlement in Donegal by investigating examples of clustered farms (clachans) and open-field cultivation—which I had been led to believe did not exist in Ireland. At the same time I was interested in the geography of the megalithic period, to which I was attracted not only by the impressive contribution the great stone monuments make to the Atlantic landscape, but also by a remark made to me by Sir Arthur Keith, then President of the British Association, when I first went to Ireland: 'I wish you would look into the question of the Ulster megaliths: I think you will find that they differ from those of the other provinces.' In the end my colleagues and I in Belfast were able to prove his point, and to demonstrate the obstinate regionalism of the Atlantic Ends of Europe, by isolating a well-defined megalithic complex which has been named after Carlingford, one of its probable places of entry. My main purpose in this address is to bring together these two lines of research and to see what light the findings of prehistoric geography can throw on the present scene.

I have called the megalithic diffusion an unresolved mystery, but it is clear that it is partly explained by the search for cultivable land along the Atlantic fringe by colonising seafaring groups from the end of the third millennium onwards. What is not yet clear is the extent to which the megalith-builders were preceded by colonists of the 'Western Neolithic culture'. Some groups, with their distinctive leathery pottery, had probably reached the Atlantic coastlands during the course of the third millennium. Radio-carbon dating is throwing a new if uncertain light on this problem, but however far back the first forest-clearances may have to be placed, it remains true that the effective exploitation of the western seaways was the achievement of the megalithic age.

Tracts of cultivable land were selected, as E. Watson has shown by detailed field study in north-east Ireland, with a keen eye for ecological conditions. Well-drained light-textured soils on the fringes of the upland areas were preferred, and the original forest cover, as evidenced both by pollen analysis and by the charcoals recovered from excavated sites, consisted mainly of oak and hazel, together with elm and alder at lower elevations and pine and birch at higher sites ranging up to the 1,000 ft. contour, which is the extreme upward limit of cultivation in the area today. In clearing the land of trees man's chief instrument was the axe—the hallmark of the pioneer farmer—and thousands of his polished stone axes, lost, discarded or deposited as ritual offerings, have been recovered by later diggers to be regarded

with superstitious awe and, more recently, with scientific curiosity. It has been assumed that any considerable forest clearance had to await the iron axe, but that there was extensive if temporary clearing in the Neolithic is now beyond doubt. We should be chary of speaking of virgin or natural forest in historic times. Further, thanks to petrological analyses and distributional studies we see the polished stone axe as a commercial asset of the Atlantic coastlands. Many varieties of their fine-grained yet tough igneous and metamorphic rocks, especially greenstones and bluestones, could be flaked into rough axes and shaped much more readily by grinding and polishing than the intractable flint of the lowlands. Quite small exposures of suitable rock are known to have been exploited, and one of my cherished field memories is the rediscovery of the exiguous exposure of porcellanite on Tievebulliagh in Co. Antrim—then unnoticed by geologists—to which I was led by following the trail of Neolithic man. Cornish greenstones, the spotted dolerite of Preseli in Pembrokeshire, the augite-granophyre of Graig Lwyd in Caernarvonshire, the metamorphosed tuffs of the Great Langdale group in Cumberland, the porcellanite of Rathlin Island and Tievebulliagh; these were some of the materials which provided axes for trade with lowland Britain as well as within the Atlantic zone. From Brittany ceremonial greenstone axes—of jadeite, diorite and chloromelanite—were traded along the Atlantic coasts and beads of callais reached the western Mediterranean. West Norway exported schist axes to Jutland, and one of the oldest known greenstone axe-factories is that on the island of Bømlo between Bergen and Stavanger. There is some evidence that bands of hunters and fishers had initiated this trade in the north, and the trail followed by the megalith-builders had been blazed along many stretches of coast by mesolithic shore dwellers exploiting the rich food resources of the warming post-glacial seas. Their kitchen-middens provide the earliest archaeological evidence of man's utilisation of the Atlantic habitat from Portugal to Scandinavia. Such culture-traits as skin-boats, fishing spears and nets and wicker fish-traps and fish-pots are but samples of traditions and techniques among the fishermen of the Atlantic coasts that go back unbroken through eighty centuries.

The far-reaching effects of forest-clearance in Neolithic times have been revealed by pollen-analytical studies in bogs and lake deposits, and recently in terrestrial soils as well. The long accepted sequence of post-glacial climatic change is being vigorously challenged as evidence accumulates of the decisive rôle played by man in promoting environmental change. H. Fairhurst, among others, has recently warned us of the dangers of accepting without question vegetation-sequences established by pollen studies as evidence of climatic change. As J. W. Watson observed for Scotland nearly twenty years ago, 'Where man first worked, peats have most formed'.[1]

The Neolithic colonisation must be regarded as a formative force in the life and landscape of the Atlantic Ends of Europe. Scientific enquiry is dispelling the romantic mists which have enshrouded the prehistoric past, and there emerge plain farming folk, initiators of the secular process of clearing the forests and coming to terms with the Atlantic environment. Not only has there been general continuity of racial stocks, of economy and tradition, but patterns of settlement and the regional differentiation of cultures have persisted down the centuries. More than once, in the west of Ireland, I have sought for and found a megalithic site, typically buried in peat, near the edge of a patch of arable land, and have been able to relate both to a local outcrop of limestone. Professor Childe has shown how the fifteen cairns of Rousay, in the Orkney Islands, occur in groups each of which corresponds to a natural unit of settlement which is still defined by a cluster of crofts, and Professor Dobby has observed the frequent association of the hamlets of Galicia with ancient sites.

[1] *Scot. Geog. Mag.*, **55** (1939), p. 153.

The Neolithic pioneers grew wheat and bere (barley) and bred cattle, sheep, goats and pigs. The quality of their material equipment was astonishingly high, for example their domestic pottery was superior in fabric and finish to that of later periods. Even if allowance is made for the postulated climatic optimum of late-Atlantic times, their colonisation of the remote fringes of north-western Europe was a remarkable achievement. I believe its success must be attributed partly to the impetus of skills acquired and experience gained in sunnier climes, and partly to the exploitation of the untapped fertility of forest soils. Fraser Darling has argued for the Scottish Highlands that the edaphic conditions of the natural woodlands were far better than either geology or climate would suggest. And if the decreasing warmth of the sub-Boreal was indeed a global phenomenon, its ecological effects on these marginal lands may well have been serious. The increment lasted well into the Bronze Age, when metals—gold, copper and tin—were added to hard stones as commercial assets of the Atlantic fringe. Legends of such a golden age have haunted the imagination of the west during many subsequent periods of difficulty. One must also make allowance for the memory of past glories carried to the west by refugees, but it is significant that the great stone monuments became in time symbols of a lost fertility, possessing magic powers of making living things multiply and endowed with such potency that practices and beliefs associated with them have lived on, sometimes in Christian guise, down to recent times. The burial chambers were regarded as the abodes of ancestral spirits waiting to be reborn, and much of the fairy lore of the west may find an explanation in the identification of the 'wee folk' with the spirits of the dead.

In discussing prehistoric environmental change we may begin with the striking fact that many megalithic sites, including such famous examples as Callernish in Lewis, became covered with peat after the monuments were erected. An explanation has been sought in sub-Atlantic climatic deterioration, but attempts to discover in the deeper bog-sections a clear and universal horizon of separation between sub-Boreal and sub-Atlantic times have been unsuccessful. It is now recognised that local variations and minor fluctuations of climate are the rule rather than the exception in the topographically complex and climatically unstable margins of north-western Europe, and that the prolonged search for precise correlations with the Baltic climatic sequence was ill-advised. It has blinded specialists to cultural interpretations of environmental change, for it now appears that dramatic deterioration of soil conditions and vegetation must be related to human activity. These landscape changes are therefore of prime interest to us as geographers, and geographers have rightly co-operated with botanists and archaeologists in studying them. At Queen's University, Belfast, the Nuffield Quaternary Research Unit (1954–8) has investigated several archaeological sites in north-east Ireland now buried under bog. A reduction in the proportion of tree pollens, particularly of *Ulmus* (elm), is the first sign of what J. Iversen has termed, in Denmark, a 'land occupation phase'.[1] This is followed by a peak of *Plantago lanceolata* pollen (ribwort plantain, a weed of cultivation) and on some sites by the appearance of *Urtica* (nettle) and of bracken spores. Regeneration of woodland is indicated by a subsequent *Corylus* (hazel) maximum, and by an increase of birch, oak and ash. At Goodland in County Antrim a blanket bog at a height of 800 ft. is shown to have developed when rushes invaded a cultivated clearing. Dr. V. B. Proudfoot, a member of the Nuffield team, has studied the soil profile and associated archaeological material and has argued that shallow cultivation by Neolithic farmers accelerated podsolisation. As iron pan formed an impermeable humus layer developed at the surface and blanket bog began to grow.

G. F. Mitchell, who, following Knud Jessen, has done outstanding work in this field in Ireland, has recently reinterpreted his pollen diagrams in terms of the

[1] *Danmarks geologiske Undersøgelse*, 2 R, Hr. 66 (1941), pp. 1–67.

effects of agriculture on the woodlands since the dawn of the Neolithic. On fifteen widely scattered bog-sites in the Irish midlands, lying between 100 and 500 ft., he has found consistent evidence of vegetational changes accompanying successive phases of agricultural activity. Again, the pollen of *Plantago lanceolata* makes its first appearance in Neolithic times concurrently with a sharp decline in elm pollen. It would seem that Neolithic man may have used *Ulmus* as an indicator of soils that suited his purpose: at any rate it appears to be a sensitive index of human activity.

It is also possible, as has been argued for Denmark, that the rain of elm pollen was seriously reduced, locally, by the stripping of its leaves and twigs for use as fodder. This practice is well known in Norway and there is some evidence that it has long been a feature of the Atlantic economy. *Ulmus carpinifolia*, which R. H. Richens suggests was brought to Britain as a forage crop in prehistoric times,[1] is still exploited for leaf fodder in north-west Spain, and both here and further north the traditional use of the leaves of winter-green shrubs such as gorse, broom and holly may perhaps be regarded as a substitute for the former deciduous leaf harvest. Gorse in particular has been widely cultivated for this purpose—and for fuel—notably in Galicia and Brittany. In Scandinavia, moreover, elm-bast, shredded and pounded, is not only still used as a strengthening fodder for milk cows but has been utilised as a substitute for bread meal in times of food shortage. Such uses of the elm could have contributed to its decline in prehistoric times. The known susceptibility of this tree to disease and presumably to climatic change must not be lost sight of in considering the marked fluctuations in its pollen record, yet it would seem, since its decline is regularly accompanied by archaeological traces of farming and by an increase in the pollens of hazel and plantain, that forest clearance by man provides the most convincing explanation.

Further, the elm virtually disappeared from Ireland during a period of renewed agricultural activity which Mitchell associates with monastic settlement in the sixth and seventh centuries A.D. This Age of the Celtic Saints, deriving much of its material culture and its inspiration once more from south-western Europe and the Mediterranean lands, provides many parallels to the Neolithic period with its megalithic missionaries. Professor Emrys Bowen has shown that sub-regions of megalithic culture—no doubt reflecting sectarian as well as environmental differences—tend to reappear in the cult-areas of individual saints, and has argued for a broad continuity of tradition along the Atlantic coasts. So well did the Church adapt itself to the tribal morality of Celtic lands, in which cattle-raiding was both a ceremonial obligation and a popular pastime, that we read of Irish saints who demanded 'a cow out of every raid'. The increase of livestock would have been an important factor in checking the regeneration of cleared woodland, and by the nineteenth century, when sheep had come to dominate the uplands, much of the landscape had become almost completely naked.

Nevertheless, the early Church encouraged arable husbandry and the growing of fruit and vegetables, and many novel crops as well as new or improved implements extended their range along the Atlantic fringe at this period. Some of these innovations may go back to the pre-Roman Iron Age, or have been borrowed from the Roman world. What concerns us is that a new adjustment to the Atlantic environment, involving the cultivation of oats, the extension of tillage and pasture into lowland forests on heavy drift-soils and a correlated expansion of the practice of transhumance, brought about an expanding economy which reached its climax in a second golden age in late Celtic times. More tolerant of acid soils and overcast skies than wheat, oats became the corn of the Atlantic coastlands, food for man and beast; and S. Erixon[2] has described its rôle in the open-hearth 'ovenless thin-bread

[1] *Watsonia*, **3** (1955), p. 152.
[2] *Folk-Liv*, **2** (1938), p. 152.

cultures' of the peasantries from Portugal to western Scandinavia. The domestic fowl, too, joined the long-domiciled pig along the Atlantic fringe in the early centuries of the Christian era.

At this time also a plough with heavy coulter and share makes its first appearance in the archaeological record, capable of cutting the sod and breaking into the mineral layer of leached soils, even if it could not effectively bring it to the surface. Here the traditional long-handled shovel presumably played its part. At any rate it was used down to recent times to bury the seed on the ridges by shovelling soil from the wide furrows. The readiness with which this method of ridge cultivation was adapted to the potato crop must be one factor in the speedy adoption of this alien food in the seventeenth century. (In addition, because of its heavy yields, it was well adapted to the small plots of Atlantic peasants.) The spade-ridge or lazy-bed, too, is thought to be of Dark Age origin, though in Ireland at least the period of its great expansion came in the eighteenth century, when mattocks, breast-ploughs and specially designed traction-ploughs were also used for constructing potato ridges. In addition to the potato, which became a staple food in Galicia as well as in Ireland, other trans-Atlantic crops which took ready root in the southern fringe were the tomato and maize.

Some archaeologists claim that the light Mediterranean plough had already reached Denmark in megalithic times, but at best pioneer farmers of the Atlantic fringe were restricted to shallow and shifting cultivation. From the Irish evidence it appears that the areas which suffered the most rapid deterioration were uplands with thin open-textured soils. For some 4–5,000 years the Boreal-Atlantic forests had built up soils whose natural fertility was being slowly reduced—as the examination of fossil soils shows—by the down-wash of minerals. With exposure and cultivation the leaching process was speeded up, and on steep slopes the forest soils were almost completely eroded and became buried in scree. The quickened run-off and redistribution of water would have affected ecological conditions over wide areas, and the growth of bogs in areas of impeded drainage was initiated or accelerated.

Typical shrubs of the secondary woodland were the hazel and the birch. It is unfortunate that the pollens of many interesting species have not been counted or are difficult to identify, but it is probable that the mountain ash, the whitethorn and gorse, as well as many familiar wild herbs and flowers, extended their range in abandoned forest clearings throughout the Neolithic and Bronze Ages. Geographers have concerned themselves with what man did to the forests: should they not also enquire what the forests did to man? Many of these shrubs, herbs and flowers, followers in man's footsteps, came to be endowed with supernatural powers and regarded as bringers of luck. The folklore of the Atlantic fringe is full of the magic of the hazel twig, of the rowan and the may tree. The creamy blossoms of the rowan and the whitethorn, the gold of the gorse, of mayflower, primrose, dandelion and buttercup have been symbolically linked with a full flow of milk and butter. To this day in many parts of Ireland the self-sown hawthorn is regarded as a fairy bush which must not be interfered with. (In Norway the rowan and the wych elm, sources of leaf-fodder, were often planted near the farms and protected from felling by taboos.)

To understand the strength of these superstitious beliefs it must be realised that before the days of large-scale hay-making it was difficult to maintain milk-cattle in condition through the winter months. They were kept warm by being housed with the family and were not turned (or often lifted) out until the May Day festival guaranteed the golden bounty of summer. The long rectangular house in various regional forms served as home and byre, and the accumulation of manure thus protected from the heavy winter rains, and also enriched with soot, was a valuable if unsavoury by-product.

The first days of May and November, the beginning of summer and winter, are not only remembered as times of gatherings and rituals designed to bring good luck, but they are the days on which, in Ireland, we are still supposed to pay our rents, once payable in cattle. These festivals of 'the Elder Faiths' marked the departure and the return of migrant livestock under the system of transhumance. Wherever along the Atlantic coasts upland areas rise above the limits of cultivation, summer hill-grazing—often largely man-made—has been used to supplement the resources of infield and outfield. Plainly in some areas, dimly in others, the influence of this threefold division can be seen in present-day landscapes and economies. In some of the outermost islands and peninsulas the old system survives little changed, and in many areas it lasted until overwhelmed by various crises about a century ago. In western Norway clustered farms associated with infield cultivation by kin-groups were widespread before the redistribution, consolidation and enclosure which followed legislation in 1857, and survivals of these 'multiple tuns' have been studied by the Oslo Institute for Comparative Research in Human Culture. We note the irregular plan of the open-field plots and of the settlements—lacking the discipline of a roadway—which is contrasted with the regular lay-out of the Swedish or Danish village. These and many other points of contrast can be paralleled by those between the Irish clachan and the English village. To illustrate the bearing of agricultural tradition on social life I draw your attention to the term 'throughother', which, originally applied to the confusion of dispersed plots under the infield system, is not unfairly extended to a mental attitude which has outlived it. Multiple tuns and clachans can also be compared in their dependence on itinerant craftsmen of many kinds—of whom travelling tinkers are the last remnants—and in the custom of electing one of their members as 'king' to act as leader and to supervise communal tasks. Such arrangements were not committed to writing: similarly the dissolution of the infield-outfield system, needing no Parliamentary Acts, is rarely recorded and has thus passed almost unnoticed by historians.

For Scotland the runrig system is relatively well known and I need only mention here the work of Dr. Arthur Geddes in Lewis and Harris and the studies of crofting communities now going on in South Uist under Dr. J. Caird of Glasgow University. Irish rundale has been investigated by Dr. D. McCourt and by Dr. P. Flatrès, who has recently published an account of the rural geography of Ireland and three other Celtic lands (Wales, Cornwall and Man). Over much of Ireland permanent enclosure was the work of improving landlords in the nineteenth century. Down to 1845, before the Great Famine had accelerated change, clachan settlements with their open-fields were well distributed around the coasts and inland on the fringes of extensive areas of rough grazing, that is in environments where the need for joint effort against the unenclosed space of sea and shore, mountain and bog was greatest. As in Norway and Scotland the evidence points to original schemes of joint ownership and communal working of the infield, supplemented by shifting cultivation of the outfield. The infield harvest was apparently divided by lot, as is the custom still with the inblown wrack on some Atlantic shores. Through time the infield was divided into plots which were allocated by lot, and today the scattered plots are permanently held. Something very similar to what we suppose to be the intermediate stage is found in north Spain, where the redistribution of open-field strips is reported to take place on November 1. In Galicia, Portugal and parts of Aquitaine one finds traces of a related system of land use and settlement, with minute subdivision among kin-groups.

Many of the islands and peninsulas of Britanny preserve the clustered farmsteads, the infields and outfields (*terres chaudes* and *terres froides*), and hopelessly fragmented infield holdings reduced to individual patches sometimes as small as a square yard. On the Cornish coasts the hamlet form of settlement persists in places though the open-fields were nearly all enclosed by the end of the seventeenth

century. In Wales the dispersion of farmsteads and enclosure were almost complete by the sixteenth century, but Dr. Emrys Jones has shown that the antecedent settlement forms were varied and also that some house clusters have persisted. Mr. G. R. J. Jones has identified examples of such nucleated settlements in North Wales with open-field communities of bondmen 'which appear to have been old-established even in the tenth century'.[1] He suggests that the bondmen were the descendants of prehistoric settlers enslaved by subsequent invaders whose economy had a strong pastoral basis.

The Irish evidence points to a similar conclusion. Here and in other parts of Europe archaeology has revealed the co-existence, at least from the early Iron Age, of single farms and clustered homesteads. On philological grounds L. R. Palmer believes (*Achaeans and Indo-Europeans*, 1955) that both single farms and collective settlements were established Indo-European institutions before 2000 B.C. Certainly the distinction between warrior-herdsmen living in isolated 'raths' and lowly cultivating peasants, presumably clustered together in clachans, is clear in the old Irish law tracts. Taking into account other prehistoric continuities it is reasonable to see, among the tributary peoples, descendants of Neolithic farmers who were absorbed into the system of Indo-European institutions from the late Bronze Age onwards, though some archaeologists would see in the pastoral Beaker-folk, with their single-grave burials, the first Indo-European overlords.

However far back the origins of these collective settlements of kin-groups should be placed, we must ask why they persisted for so long in the Atlantic Ends of Europe. The peripheral survival of archaic features can be paralleled by many examples drawn from comparative anthropology. Professor Myles Dillon has shown how in language as well as in the social order India and Ireland, at the ends of the Indo-European world, have preserved a common heritage. But environmental factors which have favoured the survival of small units of land and people and flexible systems of land use and settlement must be taken into account. It is an environment which could not be effectively utilised by more complex village communities under feudal systems such as developed in the wheat-growing plains of western Europe. Rather its limited patches of good land and extensive areas of waste in mountain and moorland, water and bog offered conditions comparable with those of the still untamed landscapes of prehistoric Europe. Archaeological research has established that small clusters of from ten to fifty houses were the characteristic form of settlement in western and central Europe in Neolithic times. We also know that the houses were normally long low rectangular structures similar in plan to those which have persisted in parts of the Atlantic fringe.

Thus many elements of the historic Atlantic culture were much more widely distributed in prehistoric times, and their occurrence in Norway indicates that some of them should be thought of as belonging to an Indo-European rather than a Celtic complex, overlying and absorbing still older coastal cultures and becoming adapted to the environment of the Atlantic fringe. Many of the concomitants of the Atlantic system—infield cultivation by kinsmen, multiple farms, partible succession, patronymics, shifting outfield patches, and transhumance—have been inferred for various parts of lowland England in Romano-British and Anglo-Saxon times. Reference should also be made to the infield-outfield system which persists sporadically in difficult environments on the European mainland, for instance in association with hamlet settlements (Drubbels) in the heathlands of north-west Germany.

In the example just quoted the outfield is burnt before cultivation, a practice widely attested in western Britain in the eighteenth century which has been regarded by economic historians as a pernicious novelty. In fact it was described at that time as a very ancient system in Ireland, and I believe it perpetuates the

[1] *Institute of British Geographers*, **19** (1953), p. 52.

prehistoric practice of clearing by burning. Rye is traditionally sown on burnt outfield both in Galicia and in Tröndelag.

If today rural settlement in the Atlantic Ends is almost completely dispersed, the past is perpetuated in other ways—in rights of common pasturage, in the small size of holdings, their rough equality and (not infrequently) their intermixed though enclosed fields. Moreover, the well-established system of ley-farming ('taking the plough round the farm') has grown out of the outfield practice of 'leaving the land to rest'. The cover of seed grasses not only checks the leaching and erosion to which land long cultivated is liable in an Atlantic climate, but is a renovator of soil fertility. And there are social continuities also. Strong ties of blood and of friendship (which in Scotland and Ireland are synonymous), nurtured in small neighbourhoods, find expression in many forms of mutual help. Traditions of friendliness and hospitality have outlived the cultural frame in which they were moulded. Impersonal external authority is suspect and the power of the modern state, geared to serve very different conditions of life, is reluctantly acknowledged. An Old World peasant ethic, aimed at restricting needs and expanding population, has contributed many millions of hard-living emigrants to the New World across the Atlantic, but despite the demographic upheavals of recent centuries, peasant values and a reverent attitude towards the land have not entirely succumbed to economic forces.

In the more remote Atlantic Ends we still have the opportunity of studying, within these islands, the impact of industrial society on traditional peasant ways of life, an impact which is world-wide in its incidence and critical in its consequences. The geographer can contribute to such studies something more than distributional and statistical surveys, for any human discipline that fails to take account of the historic factor neglects a fundamental element. On the other hand, the static reconstruction of past cultural landscapes that has come to be regarded as the aim of historical geography should be, for the geographer, a beginning and not an end.

Perhaps this is not the proper place to praise the work of the School of Scottish Studies, which has its home in the University of Edinburgh, but geographers must acclaim the spirit of co-operative research which inspires it, for it is our conviction that human beings and their environments should be studied in all their interrelations. To approach an understanding of regional variety we need to know all we can about the history—and prehistory—of human occupance, for man's choice of action at any point in time is restricted by actions previously taken. Cultural facts serve as a buffer between man and his environment, and we delude ourselves if we suppose that we need take account of the physical environment only. We should always be suspicious of simple environmental explanations such as that which relates the scattered farms of the Atlantic fringe to the ubiquity of water supplies.

The diversities of environment and society within our islands have at many stages, as Professor Fleure has reminded us, led to cross-fertilisation and the release of initiative and objectivity of mind. However much one despairs of the attempts of the Lowlanders to understand and handle the folk of the Atlantic Ends, one must hope that further movements towards the restoration of some measure of self-government will not lead to artificial isolation.

INFLATION AND THE BRITISH ECONOMY[1]

ADDRESS BY PROFESSOR A. J. BROWN

PRESIDENT OF SECTION F

WHEN I caused myself to be billed to talk on the subject of 'Inflation and the British Economy', I had not, as so often happens, given much thought to the precise interpretation which I should accord to so wide a theme. On reflection, it seems to me that I should not be wise to talk primarily of the forces which have produced inflation in this economy in the recent past, still less of the measures which might be taken to stop it. These, it seems to me, are industries in which, as in a number of others at present, there is a considerable danger of redundancy, and though I must confess to having been employed in them, it seems prudent for the present purpose to move over to a somewhat different, though related trade. I propose, therefore, to examine some leading aspects of the way in which the British economy has worked in the last ten years of peacetime inflation. This will involve some comparisons with the way in which it worked in the decidedly non-inflationary period of the nineteen-thirties. I must say, also, before going further, that there are some very important aspects of the economy's working in these years which I propose to exclude from my survey. In particular, the whole of the external trading and payments position of the British economy has been so much discussed of late that I see the shadow of redundancy there too, and I am very happy to leave it alone.

When we think of the last ten years, or of the whole period since the war, the most striking feature which marks it off from any previous period in the experience of this economy (apart from the continuous price-inflation which has gone on) is, of course, the astonishing absence of unemployment. We have enjoyed this absence long enough now for it to be perhaps worth reminding ourselves that it is, indeed, astonishing in the light of previous history, and of inter-war history in particular. We have had a decade (not to go back to the war and immediate post-war years) with an average unemployment rate of about 1½ per cent as against an average which if measured on our present statistical basis would probably have been about 11 per cent (and never substantially less than 8) in the whole period between the two world wars. We have surpassed in this respect the hopes which Lord Beveridge set out in 1944 in his book *Full Employment in a Free Society*—hopes which many other economists at the time regarded as over-sanguine. It is perhaps worth looking a little more closely at the relation between these aspirations, based upon inter-war experience, and what has happened since.

Lord Beveridge quoted Mr. Saunders' estimate that the average rate of unemployment due to seasonal variations in activity alone had, in the inter-war years, amounted to 2 per cent of the insured population, but hoped that, under pressure, this might be reduced to 1 per cent. To this he added a further 1 per cent for 'structural' unemployment connected inevitably with qualitative changes in the demand for labour in the course of technological and economic progress. To this, again, he added another 1 per cent as an allowance for those fluctuations in export demand for which it would not be practicable to compensate by manipulating the level of internal effective demand in the country. This gave him his target level of 3 per cent—about twice the rate of unemployment which we have actually experienced in the last ten years.

It may be claimed that the Beveridge estimate, and its generally less optimistic rivals, were based upon the tacit assumption that the demand for and supply of labour were to be in equilibrium in the sense that the total number of extra

[1] Address delivered on August 29, 1958, at the Glasgow Meeting of the British Association.

workpeople whom unsatisfied employers would like to take on should be equal to the number of able and willing people in the economy who are out of work—that is to say, that the excess demand for labour in the trades where it exists should be equal to the excess supply elsewhere. It may be claimed, further, that since the war there has been, not equilibrium in this sense, but a net excess of demand—hence, some people would say, our inflation. I may say in passing that I do not subscribe to this last deduction; I see no reason to suppose that, with our particular mechanism for determining wages, the general level of wages should be stable (still less that it should rise at the same rate as labour productivity) when the excess demand for labour in some trades or districts equals the excess supply in others. Leaving this aside, however, we can note that a very careful estimate has been made by Messrs. Dow and Dicks-Mireaux[1] of the amount of net excess demand in the British labour market since the war, which sets the average excess over the years 1948–56 at only one-third of 1 per cent of the insured population. If this is correct, it is clearly insufficient to explain the gap between the Beveridge target level and what has been realised.

My own suggestion is that the Beveridge target might well have been over-optimistic but for one factor which was imperfectly foreseen, and which may have played a very important part in realising the still lower level of unemployment which we have reached. That factor is a change in the attitude of employers, from regarding labour as a commodity always in perfectly elastic supply to treating it as something which, if once released, may not be easily replaced. In short, it is some degree of labour-hoarding. At what level of unemployment it became general we cannot say; that critical point no doubt occurred during the war. Nor is it easy to produce statistical evidence for its existence, but one can, I think, find one or two pointers.

Before the war employment fell off very promptly and sharply when production fell. Between 1929 and 1932, the index of industrial production fell by about 15 per cent and total real output by perhaps half as much. At the same time, the number of insured employed fell by 11 per cent. The other recorded reduction in labour-power employed—the reduction in average hours of work—seems to have amounted to some 3 or 4 per cent. Again, in the recession of 1937–8, the index of industrial production fell by about 5 per cent, and total real output in the economy by about 3 or 3½ per cent. Employment fell about 3½ per cent and hours of work by about 2 per cent. In short, before the war, a reduction in the demand for labour was met much more by reducing numbers in employment than by reducing hours of work.

Since the war, we have had only one genuine recession in production so far, that of 1951–2. On that occasion, the index of production fell by 3 per cent, and total real output by about 1 per cent. Employment seems to have fallen by about 1 per cent as well, but hours of work by nearly the same amount. More generally the year to year percentage in hours worked in the last decade has always, I think, been greater than the corresponding (opposite) changes in the unemployment percentage—this has been very clearly true in the last two years. Before the war, changes in hours worked were always small in comparison with the corresponding changes in the unemployment percentage.

The greater variation in hours than in numbers employed is, of course, natural when there is difficulty in getting more employees and when the marginal units of labour are being paid for largely at overtime rates. It may be argued that, in so far as it indicates unwillingness to lay employees off when they could, for the time being, be dispensed with, it not only amounts to a concealment rather than a genuine reduction of unemployment, but hinders progress and flexibility by keeping men out of the labour market who could be in it. It is hard to pass judgment on this

[1] Oxford Economic Papers, February 1958.

argument. The records of the last decade in progressiveness and adaptability will have to be looked at later. Meanwhile, two things can be noted. In the first place, against the tendency of labour shortage to cause the withholding of labour from the market when there is a temporary or local surplus of it we have to set the powerful incentive which it provides to the efficient use of labour in all times and places of peak demand. Secondly, when one considers that the average period of continuous employment provided by some enterprises before the war to large parts of their labour was to be measured in weeks, one can hardly feel that an increase in continuity and security is to be deplored.

Let us turn now to what has been not only a main condition of the low post-war unemployment rate, but one of the most striking features of the period in its own right—namely, the absence of fluctuations of activity in the British economy. The most important cause of this has certainly been the absence of fluctuations in the world economy as a whole. We have usually imported a large part of our booms and depressions. Between 1929 and 1932, for instance, the fall in the value of our exports of goods and services was apparently three or four times as great as the fall in our internal investment—though fortunately about half of this fall in export demand was offset by the fall in the price of our imports. But there has been nothing that could be called a fluctuation in international trade since the war. The quantum of world trade has increased every year (up to 1957) except for a minute fall from 1951 to 1952 which was mirrored by a slight dip in our real exports of goods and services. There have been disturbances from world prices, at which we shall have to look later, but real international trade, and our real exports, have followed a remarkably steady rising trend.

Have we, then, been left with no trace of a 'trade cycle' in the post-war years? I do not think this is so. We have, certainly, nothing so far that looks like a cycle in fixed investment, which, contrary to what used to be regarded as the settled order of nature, has been the steadiest part of the national product—it has risen (in real terms) every year since the war. What we have had is a ghost—or perhaps something more than a ghost—of an inventory cycle. According to the work of Dow and Dicks-Mireaux, to which I have already referred, the peak years of pressure on our labour resources have been 1947, 1951, and 1955. These have all been peak years, also of accumulation of stocks. The troughs of excess demand for labour according to the same estimates have been 1950 and 1952, which are trough years of stock accumulation also. It is true that 1957, which looks as if it may turn out to have been another peak of inventory investment, will have spoilt the run by being anything but a peak year of demand in general, though 1958 still has a chance to redeem the record by providing a coincidence of troughs. What is interesting, however, is that we have had here something much more like the inventory cycles in the United States than has been observed before. Admittedly, we know extremely little about inventory investment in this country for any earlier period; but such estimates as there are of its course in the nineteen-thirties suggest that it followed the main cycle in fixed investment.

Now, inventory cycles can be quite formidable fluctuations. In the United States the amplitude of changes in inventory investment seems to vary from 3 to 5 per cent of the gross national product, and our cycle, too, has been up to that standard. How is it that it has not disturbed the economy, and employment, more than it has? Before trying to answer this question, we should perhaps note that it under-states rather than over-states the problem, because the fluctuations of inventories in industrial and commercial hands have been supplemented rather than offset by changes in the behaviour of consumers. If we try to draw up a simple Keynesian consumption-function connecting consumers' expenditure on goods and services with the income which consumers (after paying taxes) had available for spending, we see that expenditure is for some years above the line, and for some years below.

Up to 1949, it was kept down by rationing and shortages; the first post-war restocking boom was not reinforced by anything that could be called 'abnormal' buying by consumers. In 1950, however, there was an enormous buying spree, compounded no doubt of making up for lost time and providing against a possible spreading of the Korean war. This explains, in part, why the great de-stocking of 1950 did not mean more than a slight slackening of the pressure on resources—it was a matter of increased final buying rather than reduced orders for production. But the next year—1951—brought a combination of continued high buying with very heavy accumulation of stocks; the two reinforced each other in putting pressure on productive capacity. Similarly, in the recession of 1952, not only did inventory accumulation fall to nearly nothing, but consumers' spending was also well below its 'normal' relation to available income. Wardrobes, in particular, were full, and clothing purchases fell. Similarly, 1955 was a year when abnormally heavy consumer buying supplemented stock accumulation, though the degree in which this buying was abnormal was less than it had been in the previous year. Although abnormally heavy or light final purchases sometimes offset (indeed, explain) changes in inventories, therefore, it seems that at least in the peak years of 1951 and 1955, and the recession year 1952, what we may call the inventory policy of consumers has supplemented the changes in trade inventories, leaving us with a heavier task in explaining why these fluctuations did not upset the economy more.

The main systematic reason is that there has been a very close connection between the changes in inventory investment and the changes in imports. On the average about half—and sometimes much more—of the change in inventory investment from one post-war year to the next has simply corresponded to a change in imports (if we measure both at 1948 prices). This was particularly marked in the peak years 1951 and 1955 and the trough year 1952. What it means in the present context is that only about half (sometimes less) of post-war changes in inventory investment have affected the pressure on our domestic resources; on the other hand the remaining half or more has affected the pressure on our external payments position. One might say, with only a moderate degree of over-generalisation, that our four-year inventory cycle accounts for alternate members of our series of biennial exchange crises (the rest being connected rather with events in the overseas sterling area). Inventory investment may be regarded as an exchange of monetary reserves for reserves of goods.

Apart from this reason why our fluctuations in inventory investments and consumers' buying habits have not borne more heavily on the economy, there are a number of important but less systematic factors to take into account. The first of these is the fact that the course of public expenditure on goods and services has been, on balance, such as to offset the inventory cycle. These purchases fell from 1946 to 1947, by an amount which was, in fact, much greater than the rise in inventory investment, and which allowed for substantial real expansions of fixed investment and consumption. They rose again to 1953, and have since been remarkably steady (in real terms).

The biggest real increase in a single year came between 1951 and 1952, so that it coincided with a fall in inventory investment and with slackening total pressure on our resources. It seems clear, however, that the timing of these changes in public spending in relation to the fluctuations of the economy has been much more a matter of good luck than of good management. The great fall in public spending up to 1947 was, of course, due to the clearing away of commitments left from the war; the increases of 1949 and 1952 were due to defence programmes occasioned by outside events. We cannot claim to have operated a deliberate stabilising policy through the size of government spending. Perhaps, in the absence of severe fluctuations in private fixed investment, it is hardly to be expected that we should.

But the way in which public expenditure is financed is another matter. For the

present purpose, we may perhaps confine our attention to the total true surpluses or deficits of the public authorities. Here, the most striking features have been the large, disinflationary, surpluses of the years 1948–50, and the not quite so large, inflationary, deficits of 1952–4. These have been far from perfectly anti-cyclical. The biggest change from a deflationary to an inflationary pressure came in 1950–1, when inventory investment was increasing. On the other hand, the further increase in the rate of net borrowing up to 1953, the change to a small surplus by 1950, and the resumption of borrowing in 1956 were certainly, on the whole, anti-cyclical influences. On balance, public finance can probably be claimed to have steadied the course of the economy rather than the reverse.

It remains to mention another influence which has affected the course of the economy—this time one which was outside the control of British government policy—namely, changes in world prices. So far as the level of activity in this country is concerned, the influence has certainly been a stabilising one. In the booms of 1947–8 and 1950–1, and again in a much smaller degree in 1955, the greater rise of import prices than of export prices certainly tended to drain away purchasing power and to depress profits—this is always (so far as it goes) a stabilising influence so long as our booms coincide with those in the world commodity markets. In the same way, the great fall of import prices relatively to export prices in 1951–3 no doubt had an immediately stimulating effect. This is all subject, of course, to the maintenance of reasonable steadiness in the *volume* of our exports, which was, as we have seen, a feature of this period.

But the rise of world prices—particularly import prices—has, clearly, been in another sense inflationary. While it has had a depressing or moderating effect on the level of activity, or on the excess demand for labour, it has worked powerfully, though with some time-lag, to raise prices and wages. The largest annual increase in wage rates since the war was in the recession year 1952. In discussing short-run changes, at least, it is necessary to distinguish very carefully between those things which are inflationary in the sense of tending to expand activity and the pressure on our domestic resources, and those which are inflationary in the sense of tending to raise the prices of those resources, and their products. It has been all too often assumed in the past that the two things are the same—or, at least, that they coincide.

So much, then, for the forces which have tended to keep the level of employment steady, in spite of a very considerable fluctuation in inventory investment and consumer's habits. As compared with, for instance, the United States economy, where inventory fluctuations of about the same amplitude (relatively to total product) have had more serious consequences, our greater steadiness seems to have been due partly to the greater 'openness' of our economy. Inventory fluctuations have been largely offset by variations in borrowing from, or lending to, overseas primary producing economies and, nevertheless, our exports of goods and services (largely to those economies) have been maintained fairly near to a steady trend. In addition, our financing of public expenditure has been in the main 'functional', and we have been lucky in that the variations in the public authorities' real demand for goods and services, though mainly imposed by outside events, happen to have helped stability rather than upset it. These, and no doubt other factors have enabled us to keep astonishingly free from fluctuations in our fixed investment. The element of luck in these factors should warn us against expecting their continuance.

I propose now to turn to another point, to which I have already referred; namely, the extent to which the economy in this period of full employment and inflation has shown a high or low degree of adaptability. Now, adaptability is not easy to measure, and for this purpose I have been content merely to measure, in rather a crude way, the degree in which the distribution of the occupied population between the twenty-four Orders of the Standard Industrial Classification has changed, and to compare the result with one similarly calculated for the very different period

from 1929 to 1937. The method is simply to calculate for each industrial group the percentages of the total occupied population engaged in it in the years 1948 and 1956 respectively, to take the difference between these two percentages for each group, and to add them up for all the groups. Oddly enough, the result of this calculation for 1948 and 1956 is very nearly the same as (in fact a trifle more than) for the years 1929 and 1937. By this measure, the amount of change in the structure of the economy was about the same in the recent period as in the pre-war one. There is not much else that these two periods have in common. Not only was the first a time when there was no shadow of labour-shortage in any industrial group, and the second a time through most of which every industrial group experienced some shortage, but the rates of growth of employment were different, too—the numbers in work expanded by about 11 per cent in the earlier period against 7 per cent in the later one. The identities of the relatively expanding and contracting groups were largely different in the two periods, also—in a great measure, no doubt—because one was a period of contracting exports and the other a period of trade expansion. Some groups, it is true, were expanding their share of total manpower in both—distribution, banking and insurance, gas, water and electricity supply, the vehicle industries and the food industries. Others were losing in both—agriculture, mining, textiles and clothing for instance. But the chemical, engineering and metal industries were expanding their share after the war though not before it; so were paper and printing; and the professions were expanding very much faster after than before. On the other hand we have the rapid fall in public administration and miscellaneous service since the war, in comparison with rapid increases before it. With all these differences, the total amount of change, measured in this simple way, turns out to be about the same in the two periods. But it is, perhaps, worth noting (though it is not surprising) that the similarly measured amount of net change in the rather longer, and greatly disturbed interval between them—that is to say, between 1937 and 1946—turns out to have been more than twice as great as in either of them.

It seems, on the face of it, unlikely that the process of change in the distribution of the labour-force is the same under full employment as it is with plenty of labour to spare. Before the war, there was a highly elastic supply of labour for almost any enterprise, anywhere. Now, the presumption is that an enterprise, or an industry, can expand its labour-force only by outbidding someone else. Is there any trace of this difference in the available statistics?

One way to look for this would be to see if there is any greater tendency now than there was before the war for the industries that pay the highest wages to be those which expand most. Perhaps there is, though I have not looked into this as thoroughly as I should like. Certainly there is extremely little correlation between average adult male earnings in an industry and its rate of expansion before the war; for the period 1948–56 there is some sign of one, though it is spoilt by such notable exceptions as the high earnings and decline of employment in coal mining and the rapid growth of employment in the food, drink and tobacco trades or the gas, water and electricity industries in spite of earnings rather below the average. But one would not really expect the correlation to be very good; differences between industries in degree of skill required, ease of entry and inherent attractiveness all combine to prevent it—not to mention differences in the relative strength of organisation of employers and wage-earners.

Some of these disturbing factors ought to be reduced if we ask, not about the relation between growth of employment and level of earnings in an industry, but about that between growth of employment and growth of earnings. Is there a greater or smaller tendency now than before the war for industries to increase their share of the labour-force only by raising the average earnings they provide for their employees relatively to those in other industries?

When we ask this, we come upon a paradox. For the years from 1929 to 1937, there was a definite tendency for the industrial groups which expanded relatively to the rest to be those with earnings which fell in relation to other earnings.[1] Such relatively shrinking industries as agriculture, coal-mining, precision instruments and jewellery, and metal manufacture did better than the average in weekly earnings; their relative gains in remuneration were, on the whole, more marked than were the relative losses of those other shrinking industries, such as textiles, leather and clothing, where earnings rose less than average. In the same way, the relatively expanding industrial groups—gas, water and electricity, national and local administration, building and contracting, banking and insurance—did not keep pace in earnings with the economy as a whole; distribution was the only important expanding group where earnings improved much more than the national average.

How can we explain this tendency? Before trying, one should perhaps point out that it is, in any case, only a vague one, only just sufficiently marked to appear worth explaining at all. For what it is worth, one could of course explain it by saying that the changes in labour distribution in this period must have come mainly from the supply side rather than from relative changes in the demands for different kinds of product. The high general level of unemployment makes it unplausible to suppose that spontaneous movements of labour from one industry to another made big differences to the relative scarcities of labour in different branches of activity; it seems more reasonable to think of changes in relative wages, brought about by different bargaining institutions in the various industries, as being the source of changes in the labour-supply situation. If we follow this line of argument, we can say that there was some tendency for labour in the industries where it secured the biggest wage-increases to price itself, or its products, out of the market.

But that is not the only possible line of explanation, or even perhaps the most likely. In so far as differences in wage-fixing institutions were dominant in determining the relative alterations in wage-rates in the thirties, it may be that the negative correlation between wage-changes and employment-changes arose from a general tendency for trade unions to be strongest in the static or declining industries. One can think of reasons for a tendency of this kind. Labour is perhaps less likely to be organised in an industry which is recruiting workpeople fast than in one where the labour force has been long established. Industrial expansion between the wars was largely concentrated in parts of the country where trade unionism was, initially, relatively weak—an association which was only partly a matter of cause and effect. Strong labour organisation may well be, on the whole, an attribute of mature industries.

But whatever may have been the reason for the pre-war connection between relative growth of employment and relative decline of wages and earnings, it is clear that this connection has subsequently ceased to exist. Mr. Nicholson[2] showed that there was, in general, an association between high increase in employment and high increase in wages in the years 1938 to 1943, and it seems that the same kind of association has existed in the period from 1948 to 1956.

For this post-war period the correlation between changes in economy and changes in employment is certainly positive—if the enquiry is confined to adult male wage-earners, it is quite high.[3] Leather, railways, clothing, agriculture and public administration are conspicuous for relatively low increases in earnings, and low (or negative) rates of growth; printing, chemicals, metal manufacture, and food, drink and tobacco for high growth both of earnings and of employment. The vehicle industries, with one of the highest rates of growth in employment, show a growth

[1] The correlation coefficient is $-0{\cdot}41$, which stands near the 5 per cent level of significance.
[2] Oxford Institute of Statistics, Bulletin, 1946.
[3] Mr. R. Wilkinson has calculated a correlation of $+0{\cdot}22$ using total labour employed and its weekly earnings, but $+0{\cdot}61$ using adult male labour and its hourly earnings.

in hourly earnings which is less than the average, but there are no other large departures from the general tendency.

It seems, therefore, that while the economy has changed its structure about as fast under full employment as it did in the pre-war period of high unemployment, there is, indeed, some sign that the growing parts have had to outbid the rest for labour; it has not come to them (as it did before the war) in spite of a relative fall in the remuneration they offered it. This change has, no doubt, some significance in connection with the mechanism of inflation. When wage-rates rose generally before the war (as they did in the middle and later thirties), the rise was not distributed according to the changes in their demand for labour; rather it seems to have been a matter of trade union organisation and enterprise. No doubt these qualities are still very important, but the different extents to which they have been helped by different strengths of demand for labour as between industries have emerged as a major feature of the scene. Perhaps one can say, too, that full employment has tended to make the bargaining-powers of labour in different industries more nearly equal than they were before the war—the proportional relative changes in earnings between the industrial groups have, in fact, been rather smaller in 1948–56 than it was in 1929–37.

So much for the flexibility of the economy under inflation. It is appropriate next to look at its record of progress. It seems to be very widely accepted that this has been good in relation to previous times, if not specially brilliant in comparison with other economies (also undergoing inflation) in recent years. The rate of growth of real domestic product has, since the war, been something like 3 per cent a year, which is about the same as the rate achieved in late Victorian times, and a good deal faster than we have managed over any considerable period in between. But whereas fully half of the Victorian expansion was directly due to the growth of the labour force, our population in work has been growing at only half the Victorian rate, and output per person in work has grown by about 2½ per cent a year.

This is a respectable, but not, by modern standards, a sensational rate. It is misleading to compare it with the German expansion of recent years, which has contained a stronger temporary element of restoration, as contrasted with genuine growth. It is also misleading, for some purposes, to compare it with the contemporary growth of Russian real income (probably about twice as fast as ours), which is made possible in part by the application of techniques as modern as our newest in place of some which we have long outgrown. But it is more to the point to note that our real output per head since 1948 has still been rising less fast, though not so very much less fast, than that of the United States.

It is natural to connect our rate of progress with our rate of net investment. We have accumulated capital faster (on the average) than before the war—which can be associated with our higher rate of growth as compared with then—but also much more slowly, in relation to the size of our output, than the Russians or the Americans —hence, one might say, our lower rate of growth as compared with their economies. Capital accumulation and growth of output certainly go together, but the connection between them is very far from simple. How far it is from simplicity can be seen from two aspects of our recent experience. The first is that the temporal relation between investment and the growth of output per man-hour, in the short run, is not very clear. There seem to have been two periods of rapid increase in physical output per man-hour since the war; one which brought a sharp peak in 1949, and another perhaps rather more sustained but less rapid in 1953 and 1954. These spurts certainly do not follow bursts of rapid net investment, as one might expect. The first post-war burst of net investment, whether one considers all kinds of accumulation, or all excluding dwellings, or simply net investment in ships, vehicles, aircraft, plant and machinery, was only just getting into its stride in 1949, and increased in intensity for a year or two after the rate of growth of output per man-

hour had started to fall. Then, again, the much bigger investment boom which we have recently been enjoying was very far from over when the rate of increase in labour productivity passed its latest peak in 1954. We shall certainly not make much sense of that fashionable entity, the marginal capital–output ratio, from data such as these.

There is more to be said for looking at the period as a whole, or rather, for taking the interval between 1948 and 1955, two years in which the degree of labour shortage seems to have reached peaks of about the same height. It simplifies matters, too, if we confine our attention to manufacturing industry, which was responsible for about three-fifths of the real increase in our domestic product over those years. Manufacturing industry increased its labour-force in employment by about 13½ per cent in that time. How much did it increase its fixed capital? By a rather doubtful calculation based upon Mr. Barna's estimates of the replacement value of fixed capital, and the Central Statistical Office's figures of net investment, it seems that the proportionate increase may well have been about the same as that in the labour-force—about 13 or 14 per cent. But the output of manufacturing industry increased by 42 per cent. That means that it increased in relation to both the labour and the capital employed by some 25 per cent—about 3 per cent a year. This is increasing returns with a vengeance. What the experience suggests is that more of the increase in manufactured output comes from improvements in the quality of capital—from technical progress in the strict sense—and from improved supply of other factors, such as managerial skills, than from sheer capital accumulation. Or one may say that, since manufacturing capital seems to have increased no more than in proportion to the amount of labour working with it, the very substantial increase in output per man-hour is to be attributed not to increase in the quantity of capital (as it might be measured by its cost of production at constant factor-prices) but entirely to improvement in its quality and to such improvements as there have been in supplies of other factors. This no doubt helps to explain why the connection between net investment and improvement in output per man-hour in the economy as a whole has been so loose. Its looseness, however, in no way detracts from the importance of capital accumulation and capital replacement as the means by which improved techniques are brought into practice.

This brings us to a final subject, which has, perhaps, more direct connection with inflation as such (as opposed to full employment) than those which I have been considering. How has the economy reacted (especially in its spending and saving habits) to the increase of prices? Has the expectation of inflation become an accepted part of life? It has often been pointed out that there are certain difficulties in envisaging a long-lasting and general expectation of continued inflation. Where prices are flexible, and are formed in free markets, such an expectation will at once generate a price-increase of just such a size that the general expectation of further increase is itself extinguished. If in some way it is continuously regenerated, then the only possible outcome is hyper-inflation. A lasting and generally-held expectation of inflation, therefore, in an economy which has obviously not suffered hyper-inflation, can exist only in those markets where prices are not formed as the more or less instantaneous outcome of expectations. Now, there are many markets where this is the case. Indeed, it is only in the primary commodity markets (if we exclude purely financial markets for the moment) that prices are completely sensitive to expectations about their future. Even there, also, we have to make an important reservation. A general expectation of price-increase proceeding at an annual rate less than the percentage annual cost of holding stocks of the commodity will not cause the current price to rise. Since the war, the course of primary commodity prices has fluctuated; there have been two major decreases to discourage expectations of anything like continuous increase; and the rate of ascent of the trend about which the fluctuation has taken place must in most, if not all, cases have been less

than the percentage cost of holding the commodity concerned in stock. The existence of widespread expectations that prices will be higher in future in nearly all markets is therefore a very decided possibility.

What effect would such expectations have? The main effect, of course, should be to divert preferences away from money and assets of fixed money value, towards assets of fixed real value and towards immediate consumption. It is worth looking to see whether there are any signs of these shifts.

I have already referred to some of the features of a simple Keynesian consumption-function relating to the personal sector of the economy in post-war years. If such a schedule is drawn up with disposable income and consumers' expenditure measured at 1948 prices, and the data for a pre-war year are inserted, it becomes plain that the schedule may have shifted upwards a little since before the war, but not by very much—consumption may be £200 or £300 million higher at any level of total real disposable income than it was before the war. Does this indicate a reduced willingness to save, because of fears of inflation? Not necessarily. One has to remember that there has been a very substantial shift in the distribution of disposable personal incomes since the war; the million of the total population with the largest incomes now have little more than half the percentage of aggregate personal income which the richest million had before the war. Again, the evidence is that, in the long run, as average real income rises, the proportion of a given real income which is saved tends to decrease, perhaps for reasons which have been set out by Professor Duesenberry, perhaps for the simpler reason that economic progress seems to consist largely in the invention of ever more seductive kinds of consumers' goods. One would expect some shift in the schedule after a decade or two. Some shift due to apprehension of inflation is not excluded, but I do not see any positive sign of it, and it cannot be very large.

One might expect to get more definite signs of any general change in expectations from the public's attitude towards different kinds of asset. But this is not a simple matter either. The relative supplies of the different kinds of asset have changed very greatly since before the war, and the drastic increase in taxation of yields must have had some effect on behaviour also. For the last decade as a whole, the yields of equities have been above those of bonds by about the same amount as in the nineteen-thirties; the gap has been greater than in the periods of recovery and prosperity before the war. Other things being equal, this might be taken to mean that preference for fixed interest assets is somewhat greater than in the rising phase of a pre-war trade-cycle. Among the other things that are not equal, one—the greatly increased supply of bonds in comparison with equities—may seem to support this suggestion. More bonds in relation to equities should, by itself, mean cheaper bonds in relation to equities, and thus a bigger gap between their respective yields, and this, as we have noted, is contrary to what we find. Perhaps the other big change in the situation works the other way; assets providing uncertain money incomes, as equities do, are notoriously rendered less attractive by high marginal rates of tax. The comparison of bond and equity yields, therefore, gives no certain indication of any change in the public's attitude; but, again, it seems to me to show that there has at any rate been no large shift of preference away from fixed interest securities.

The other question which one is tempted to ask in this connection is whether the schedule of liquidity preference has shifted. A general expectation of a fall in the value of money—even a slow one—might be supposed to reduce the amount of money which the public would be willing to hold at any given level of interest-rates—presumably net interest-rates after the deduction of an allowance for taxation. To reach a firm conclusion about this, is, again, not easy. It is possible to make allowances not only for taxation and for the changing requirements of money for transactions purposes, but also for the supply of liquid interest-bearing assets in

the hands of the public, and for the level of dividends on equities. When all this is done, there seems to me to be nothing about the present situation which is out of line with general experience in the last thirty years.

I have tried in this lecture to range over a number of aspects of the internal working of the economy in the period of inflation and full employment since the war. Perhaps the chief general conclusion that emerges is that the difference that price-increase and absence of unemployment have made to it are remarkably small. I see no sign that expectation of continued price-increase has become engrained in the system; there are spasmodic signs of it in the distinct separate booms of the post-war years, but for the period as a whole it seems to be notably absent. There is no sign that saving has been seriously affected by inflation, and by achieving fuller employment and higher real income than before the war, we have increased the proportion of the national product devoted to investment; we have also achieved an accelerated rate of growth, connected with improved technology even more than with the enlargement of the capital stock. The occupational and industrial structure of the economy has apparently been about as flexible under full employment as it was with a great deal of unemployment, though the price-mechanism has been more patently the servant of demand changes in bringing it about.

The key to the positive achievements of this period has been full employment. This discussion suggests that it may be very easy to assume too much in taking the continuance of this happy state for granted. We ought perhaps to pay more attention to the peril to stability inherent in an inventory cycle; we have weathered very considerable fluctuations of this kind fairly smoothly with the help of an appreciable amount of good luck (in the form of fortunately-timed changes in defence expenditure) to supplement our not altogether bad management. But the far greater danger, which we have so far escaped, is that from a depression in world demand. Our weapons for combating this are untried; it is not even certain that they exist. And, lest it should be thought that any considerable lapse from full employment would at least relieve us of the embarrassments of inflation, I feel obliged to say that this, too, is unproved. My own view, indeed, is that one of the less welcome marvels of modern times is the possibility, within limits, of enjoying inflation and depression at the same time. But to discuss that would take me beyond the task I have set myself for this occasion.

HEAT AND THE ENGINEER[1]

ADDRESS BY PROFESSOR O. A. SAUNDERS, F.R.S.
PRESIDENT OF SECTION G

DECIDING ON THE SUBJECT of this Presidential Address gave me some difficulty, since for one who has spent most of his life in education there was a strong temptation to speak on the many problems which confront us in teaching engineering today, about which most of us have ideas and are only too willing to express them. But I should not care to risk comparison with the inspiring address by your last year's President, Sir David Anderson, on this subject. There was also some attraction towards such topics as gas turbines and nuclear power plant with which I have been connected from time to time, and which perhaps have a more popular appeal than the topic I have chosen, but in the end it seemed that the subject of heat, and particularly heat transfer, is of such wide importance in engineering today that it offered the best scope, especially since it is the subject on which most of my own scientific work has been centred.

THE IMPORTANCE OF HEAT AND HEAT TRANSFER IN ENGINEERING

I like Thomas Tredgold's definition of engineering as the 'art of directing the resources of nature to the use and convenience of man'. Pursuing this art the engineer, whatever his speciality, will always encounter problems of heat and heat transfer which extend from one end of the field to the other. Apart from the fact that we still know of no practicable way of obtaining power from fuels without first passing through the form of heat, the vast majority of physical and chemical processes used in industry are heat-controlled or heat influenced, as indeed are many natural phenomena including life itself. The engineer meets heat problems in both large industrial plant and small domestic appliances. They occur for example in the steam power plant and the steam laundry, the industrial furnace and the domestic cooker, the cooling of the missile or satellite and the business of keeping himself cool or warm in his home, the commercial evaporator or distiller and the ladies' hair-dryer. Only too frequently heat is in the wrong place and arrangements have to be made to transfer it to somewhere else, often at great effort and expense.

The energy stored in the Earth is found as chemical energy of fuels, or nuclear energy, and while it is not very difficult to convert it into heat, it is rarely then where we want it, with the notable exception of the internal combustion engine. Some of the most costly parts of Calder Hall are the enormous heat exchangers needed to transfer the heat from the carbon dioxide to the steam, and indeed the reactor itself is largely a heat exchanger and many of the most difficult problems of reactor design arise from this fact. And in all appliances in which temperature has to be regulated by the inflow or outflow of heat, complicated heat transfer problems arise, the successful solution of which is often vital to the performance of the plant.

THE NATURE OF HEAT

It is perhaps surprising that such a familiar concept as heat should still offer any difficulty of definition to the engineer. The history of thermodynamics and the controversies of the last century no longer greatly excite the engineer, but the importance of absolutely clear understanding of terms used cannot be overemphasised, especially with so many new developments coming along, and I was

[1] Address delivered on August 29, 1958, at the Glasgow Meeting of the British Association.

surprised to find when sitting recently on a board of advisers for the appointment of a university reader in applied thermodynamics, that apparently very suitable candidates could not answer satisfactorily the simple question 'What is Heat?' The best textbooks and especially the great writers on thermodynamics, such as Gibb and Planck, are of course crystal clear on the subject, but the engineering student of today is often loosely taught to regard heat as a property of a system, thereby confusing it with temperature and internal energy. The difficulty disappears immediately if we consider heat as a transient process of flow between systems of different temperature. It is also rather dangerous to teach the subject in terms of molecular energies and velocities unless the kinetic theory of gases is properly developed including its statistical nature. There is however no real difficulty in presenting formal thermodynamics to engineering students, any more than in presenting mechanics. These matters are not merely of academic interest when we have such new developments as electronic heat engines, molecular flow of rarefied gases, and so forth. Inventors, and even reputable scientists, still appear in my department at the frequency of one or two per year, claiming to have circumvented the second law of thermodynamics, but the inevitable conclusion is that the inventors themselves do not understand the law.

Heat and Power

The struggle to increase the efficiency of conversion of heat into power by conventional methods seems to have reached the stage when further progress is difficult and costly. To discuss the prospects of steam and gas turbines and other kinds of power plant would be a subject in itself, but it is interesting to observe how often progress depends in the end upon the solution of problems of heat transfer. For example, in steam power plant one of the crucial problems is the design of the superheater to give satisfactory heat transfer without burning out the tubes. In the gas turbine, leaving out aircraft applications, the fuel consumption could easily be reduced from its relatively extravagant figure in the simple gas turbine to a value not far off that of the Diesel engine if a really satisfactory heat exchanger could be developed. And in the nuclear reactor, as already mentioned, one of the main engineering problems is to achieve a sufficiently high coolant leaving temperature for a good thermodynamic efficiency without excessive heating of the fissionable material.

Before leaving this subject it is perhaps worth referring to the possibility of some form of direct conversion of chemical or nuclear energy into power. The fuel cell, which does this for chemical energy, has been known for many years, but the main difficulty is the very low output from a given size of apparatus, which renders it impracticable for all but a few special applications. This appears to be fundamentally linked with the small temperature changes involved. Another recent development which has attracted interest is the possibility of converting heat into electricity by some form of electrical apparatus in which a flow of electrons can be intensified by the application of heat. The American General Electric Company have experimented with a thermionic converter in which an electric flow between a hot plate and cooler one abstracts heat from the hot plate. This would eliminate the necessity of the elaborate turbo-generators and heat exchangers needed in a conventional plant, but so far the device appears to be possible only on a very small scale, suited, for example, for providing current for radar sets in remote districts. At M.I.T. they have recently reported success with a device of this kind with an efficiency of 13%. Both of these are heat engines in which the working medium is an electron gas and the work is abstracted as electrical energy when the electrons move in an adverse electrostatic field. Even if the efficiency of such apparatus were not more than 20 per cent they would still have applications because of the elimination of costly turbines and generators; it is, however, to be remembered that on a

larger scale the getting rid of the remaining 70 to 80 per cent as waste heat may be a formidable problem.

Still more interesting is the conversion of energy from nuclear sources directly into electricity, on which very little is yet known. For nuclear energy produced by fission as in existing nuclear power plant, the problem looks insoluble because the bulk of the energy is produced as kinetic energy of uncharged neutrons which appear as heat when the particles are reduced to rest in the surrounding material. On the other hand if a fusion type of process can be developed by some form of electrical discharge, energy may be produced in the form of directional motion of charged particles and it might be possible to extract some of this in a magnetic field as electrical power. Indeed this may be the only way in which such energy could be abstracted, since the problem of transferring heat at a temperature of hundreds of thousands of degrees to a solid surface without vaporising it may be extremely formidable.

The Fundamentals of Heat Transfer

The science of heat transfer remains based on the three elementary processes of radiation, conduction and convection, but to these have been added heat transfer with change of phase, as in evaporation and condensation, and with chemical reaction. The greatest difficulties of the subject arise from the complicated conditions in which these processes operate in practice. The basic differential equations of conduction and convection have been known for over a hundred years, but the boundary conditions for which solutions existed were for a long time so restrictive as to be of very limited use to the engineer. During the last quarter of a century the situation has changed greatly with the advent of numerical methods, including relaxation techniques and of digital and analogue computers. In the field of conduction today, for example, even for unsteady state problems, there is no reason why any problem however complicated the boundary conditions, cannot be solved, given the necessary time and equipment. As an example, the temperature distribution in an axisymmetrical steam turbine rotor during the initial starting up period has recently been found in my department by using an electrical resistance analogue computer, this investigation being a necessary prelude to studying the thermal stresses involved, which are crucial matters to the designer. The shape of the rotor is very complicated and indeed it is from the shape that the problem of stress concentration arises.

In convection, which is perhaps the most difficult branch of heat transfer, great progress has been made in solving the boundary layer equations for laminar flow for both free and forced convection and, as in the case of conduction, we are now in the position that with the aid of computers we can solve almost any problem. For example, solutions have been given for forced convection laminar boundary layer flow with temperature-varying physical properties, with and without fluid injection or removal at the surface. Recently cases of non-uniform surface temperature have been treated, starting by the stepped temperature distribution, which leads to the case of any arbitrary temperature distribution at the boundary. In free convection also, many new cases such as uniform heat input to a vertical plate have recently been solved. In fact, to sum up, there is not much research interest left in solving laminar flow convection problems since it is now largely a question of feeding the particular conditions into a computer.

We are left, however, with turbulence, and here the position is quite different, as indeed is the case in any branch of fluid mechanics. Unfortunately the pace of modern industry is such that the flow is turbulent in the majority of applications. We still have to rely on experiment for finding turbulent heat transfer, even though considerable progress has been made towards understanding the temperature distributions in fluids in turbulent motion. The fundamentals of the relation

between heat transfer and friction for turbulent flow were given in the middle of the last century by Osborne Reynolds, and we still use developments of his method today. An example of this is the gas-cooled nuclear reactor in which the need for minimum pressure loss in the carbon dioxide flow is very important since some 4 per cent of the final power output has to be fed back to drive the circulating fans, reducing the overall efficiency from 25 per cent to 21 per cent. The particular geometrical conditions of this application are a little unusual in that the gas flows through a long annular passage with heat flowing into it from a finned surface on the inner side, the heat flow being distributed in a non-uniform manner controlled by the neutron flux distribution, a set of conditions obviously not to be found in the text-books. The flow is turbulent and the problem is to design the fins on the inner core so as to give satisfactory heat transfer coefficients everywhere with a minimum of pressure loss. This problem is actually a new and complicated variation of a very old engineering heat transfer problem, namely to get heat transfer without excessive pressure loss by varying the geometry of the boundary such as by roughening, grooving or other devices, a topic on which large numbers of patents have been taken out for boilers, heat exchangers, etc. It has always seemed to me that, in very oversimplified terms, success depends on creating turbulent eddies in the places where the temperature gradients are a maximum and so far as possible to produce larger and slower-moving eddies rather than smaller faster ones, although, of course, there are limits to the extent to which this can be done. The developments in design of fins for the uranium cans in the new power stations will be followed with much interest.

Dimensional Analysis

Some thirty years ago the engineer received much help in correlating experimental heat transfer data, particularly for turbulent convection, by the use of dimensionless groups of variables, and the use of Reynolds, Prandtl, Grashof and Nusselt numbers and Mach numbers is now universal, with more recent additions such as the Schmidt number for evaporation, Knudsen number for rarefied gases, and so on. It is rather remarkable that only during the last few months a new extension of dimensional analysis has come to light, namely vectional dimensional analysis which is based on the consideration of vectorial lengths in, say, the three directions of Cartesian coordinates, as independent fundamental units, instead of using only one unit of length as in ordinary dimensional analysis. This immediately leads, for laminar flow, to the relation $\mathrm{Nu}/\sqrt{(\mathrm{Re})} = f(\mathrm{Pr})$ for forced convection and to $\mathrm{Nu}/\sqrt{(\mathrm{Gr})} = F(\mathrm{Pr})$ for free convection, with an interesting extension to cases of combined forced and free convection by including an additional independent variable group $\mathrm{Re}^2/\mathrm{Gr}$ on the right-hand side.

It may be objected by the mathematically minded that these results follow immediately if we write the basic boundary layer differential equations in non-dimensional form, which is indeed true, but the approach through dimensional analysis is of fundamental and educational interest, and may be of value in studying new problems. It is also of interest that vectorial dimensional analysis applies only to laminar flow, and not to turbulent flow; in fact, it distinguishes the two on purely dimensional arguments. The reason, of course, is that in turbulent flow the directions of local viscous shearing and local conduction are not exclusively parallel to the surface, as they are for laminar boundary layer flow, with the result that prescribed vector dimensions cannot then be attached to the viscosity and thermal conductivity.

Turning to radiation, apart from determination of surface emissivities for different wavelengths, the main unknown fields are semi-transparent media, and radiation from clouds of particles in the important practical case of the luminous flame. The real difficulty in all these problems lies in determining the local radiation

and absorbing the properties, which in the case of flames are linked with the rate carbon formation in the combustion process. These and other fundamental problems will be referred to again later with reference to some practical applications.

In the general field of heat transfer with change of phase, progress has been made recently with the solution of boundary layer problems with simultaneous heat and mass transfer and we are now able to treat theoretically such matters as combined evaporation and heat transfer from wet or subliming surfaces.

Since this is an Engineering Section I propose to devote the remainder of this address to some of the practical applications which involve heat flow problems.

The Regenerator

The regenerator is one of the oldest forms of heat exchanger, large brickwork regenerators beneath steel-melting and other types of furnaces being very common from the earliest days of furnaces, for the purposes of preheating air or gas to obtain higher flame temperatures and to recover waste heat. These regenerators also had the advantage that by the use of ceramics rather than metal, as in recuperators, high working temperatures were possible before the discovery of heat-resisting metal alloys.

The regenerator is, however, apart from the use of ceramics, fundamentally a very attractive principle of heat exchange because it uses the same passage for alternate flow of hot gas and cold air, instead of two passages separated by a dividing wall as in the recuperator. This eliminates the mechanical problem of separately sealing off each individual tube as in the tubular recuperator, and allows a better shape of flow passage for both fluids instead of, for example, the tortuous flow across baffled tube banks of a tubular recuperator. Also there is the possibility with regenerators of making the individual flow passages so small that the flow is laminar, and thereby reducing the ratio of friction pressure loss to heat transfer, which would be impossible with tubular recuperators in which tubes smaller than one-quarter of an inch have not been contemplated for any large-scale exchangers because of the enormous problem of sealing tens of thousands of tubes into end-plates. The thermal recovery of a regenerator can also be raised to very high values, over 90 per cent, by so arranging the temperature distribution within the solid that it resembles that in a counterflow recuperator, and by choosing the periodic time between reversals small enough for the solid temperature to fluctuate only slightly. These ideas on regenerators, which represent a new departure from the old massive brickwork regenerators of the steelworks, were first pointed out by Ritz after the war, but despite the advantages just described the use of regenerators in industry has not yet extended much because of practical difficulties in finding a satisfactory mechanical arrangement. Two methods are possible, either to use stationary solid matrices with reversing valves, or to use a form of continuously rotating matrix such as a disk or cylinder, part of which is situated in the hot stream and part in the cold, so that the rotating matrix passes from one to the other. The former arrangement is clumsy, and the latter involves some form of rubbing or rolling seal. For applications where compactness is essential, such as the vehicle gas turbine, the rotating regenerator is very attractive, but the problem of developing a seal which is effective with pressure differences of two or more atmospheres and does not wear unduly, has not yet been solved. The whole future of the gas turbine for motor vehicles, and possibly other mobile applications, would be transformed by the development of a satisfactory regenerative heat exchanger, and although there are rumours of some success in U.S.A. by General Motors nothing has yet emerged.

High-Speed Heat Transfer and Missile Heating

The principles of aerodynamic heating are now fairly well understood. The dissipation of energy by viscous action in the rapidly shearing boundary layer air

develops heat which is removed into the main air-stream by conduction or eddy diffusion, but the two processes are so related that an equilibrium temperature is reached which, if the surface is unheated, is termed the adiabatic wall temperature and exceeds the main stream temperature by an amount depending on the Mach and Prandtl numbers of the stream. This excess temperature rises with the square of the Mach number, and at $M = 5$ reaches about 1100° C for air at sea-level temperature. Aircraft flying at supersonic speeds, and still more missiles, therefore reach high temperatures even in the upper colder layers of the atmosphere, and metal models of conical missile heads have melted in a wind tunnel at $M = 6{\cdot}9$ after only about 20 seconds. Even if the surface does not melt, except when the duration of high-speed flight is short, the penetration of heat into the structure is unacceptable, and even for short durations severe thermal stresses may be caused. Various ingenious means have been proposed to overcome these difficulties. For continuous operation a source of cold is needed and in an aircraft this means either a refrigeration plant or using the cold capacity of the fuel, which is substantial especially for supersonic transport aircraft which requires large jet engines. The attractions of using a cold fuel such as liquid hydrogen are considerable since most of the sensible 'cold' is in the form of specific heat rather than latent heat. The distribution of the 'cold' throughout the aircraft structure and cabin is, however, no small problem. Another method, particularly for missiles, is to allow the metal surface, or some material coating it, to sublime into the air, thus abstracting latent heat at the surface. For all these purposes the designer needs to know heat transfer coefficients at very high Mach numbers, and although there is good evidence that the heat transfer is proportional to the difference between the surface wall temperature and the equilibrium adiabatic wall temperature up to Mach numbers of 4 or 5, there is still uncertainty concerning the heat transfer coefficients at very high Mach numbers. One difficulty is that there is evidence that the condition of flow in the boundary layer may be affected by the Mach number and also by the abstraction of heat, though this probably means a lower heat transfer coefficient than would otherwise be expected.

While on the subject of aircraft, the aeroplane is a good example of heat always being in the wrong place and requiring the heat transfer engineer to get it to the right places. First, there is abundant heat in the engine, and the engine designer is at pains to keep it away from vital stressed parts such as the turbine blades and bearings, and the waste heat in the exhaust is enormous. But the cabin requires heating and the wings and air intakes de-icing, and the designer even goes so far as to use electrical energy, converted from fuel heat at poor efficiency, for these purposes. Again the oil and fuel must be heated to prevent undue frosting or freezing. Aerodynamic heating is a further nuisance at high speeds.

In connection with aircraft and aero engines I encountered an interesting heat transfer problem in the early days of the last war, when I was asked by Sir Henry Tizard to consider means of injecting oxygen into Merlin aero engines to restore the loss of power above the rated altitude, above which the supercharger is unable to maintain the maximum supercharge pressure, the aim being to raise the maximum ceiling to which fighter aircraft could climb, then about 36,000 feet. It was immediately evident that the oxygen had to be carried as liquid, and a system was devised for injecting liquid oxygen into the air intake of the engine. A serious difficulty, however, was that on turning on the liquid oxygen from the supply tank, the 14-foot pipe leading to the injection nozzle became vapour locked and a minute or so elapsed before the steady supply could be established, which was useless from the pilot's point of view. Insulation of the pipe was no help. In the end a double wall self-refrigerating pipe was devised with a number of small holes in the wall of the inner tube through which the liquid oxygen escaped into the annular space between the two tubes and cooled it mainly by evaporation. By suitably arranging

the disposition of the small holes, the inner tube could be made to cool itself down to the boiling point rapidly enough for the liquid oxygen flow in the inner tube to become regular and full in a few seconds after turning on. Eventually the appliance was installed in a fighter aircraft and raised its ceiling from 36,000 to 38,000 feet, also improving the mean rate of climb between 30,000 feet to 35,000 feet from 650 feet a minute to 900 feet a minute with a total duration of about 15 minutes. But for the arrival of the gas turbine aircraft engines the device might well have had a much wider use.

Dropwise Condensation

For a long time it has been known that the heat transfer between a vapour and a surface upon which it is condensing, is much greater when the surface is not fully wetted and the vapour condenses as drops rather than as a continuous film. This conclusion was based on a simple measurement of steam condensing on a small surface, and coefficients of tens of thousands in British Units have been observed, compared with values of the order 1,000 for film condensation. Incidentally the precise physical mechanism of dropwise condensation is still uncertain, and we do not really know why it gives such high values. Although the absence of the liquid film through which the heat of condensation must pass to the surface obviously reduces the resistance to heat transfer, the proportion of the surface in contact with the drops is not large, and the drops themselves are not so small as to offer no resistance to heat conduction. It is disputed whether or not condensation can take place on the bare surface between the drops, and theories have been suggested which involve rapid turbulence in the vapour close to the drop surface caused by movement of the drop surface due to surface tension variation with temperature. Whatever the explanation may be there is no doubt of the fact, and the question of practical application has caused much interest especially recently. The difficulty of promoting and maintaining the dropwise condensation has been substantially overcome in the case of steam condensers by the development at the Mechanical Engineering Research Laboratory of chemical additives of the mercaptan type, very small quantities of which are found to have the desired effect for quite long periods. The practical gain is, however, as so often happens in engineering, much reduced when we come to an actual plant. This is partly because most of the heat transfer resistance is on the cooling water side, and raising cooling water velocities, with familiar corrosion troubles, is necessary to take full advantage of dropwise condensation. Other important matters are the effect of small quantities of non-condensable air which appears to be more detrimental for dropwise than for film-wise condensation, and the fact that in parts of a condenser where the steam velocity is several feet a second the steam flow breaks up the water film and produces something resembling drops even under normal condenser operating conditions. Nevertheless, it is probable that power station condensers could be something like halved in size, and perhaps in cost, if a satisfactory dropwise condensing condition could be relied upon.

For vapours other than steam, in chemical engineering processes, the attractions of dropwise condensation are considerable, but unfortunately chemists have not so far been able to find a promoting substance which will adhere to stainless steel in the same way as mercaptans adhere to a copper-based alloy.

Effusion Cooling

The cooling of a surface by forcing gas or liquid coolant through pores or slits or holes was first mentioned by Duwez. The main cooling effect arises from the insulating layer of cool fluid interposed between hot stream and the surface, though heat is also removed from the wall by the coolant direct. The effect is increased if the fluid is liquid and vaporises on emerging from the wall surface.

The theory of this process has been worked out for continuous injection, but in practice only the use of porous walls gives a continuous film, the effects of injection from slots or holes being localised. Porous injection also tends to be wasteful of coolant in the downstream parts unless the distribution of injected flow can be made non-uniform by controlling the porosity or by some other means. The most obvious applications of effusion cooling are to gas turbine blades and combustion chambers, rocket chambers, missiles, and other cases where the protecting of the surface from a hot gas enables a cheaper material or high gas temperature to be used, but it is essential to have a supply of coolant and one criterion by which effusion cooling is judged against other methods is the amount of coolant needed. It is now generally agreed that for most of the applications mentioned effusion cooling through porous walls such as sintered metals is more economical of cooling air than other methods such as internal blade air passages, especially if the distribution of coolant can be efficiently arranged, but the main reason why the method has not been used much so far is the decreased mechanical strength, both tensile and shock, of sintered metals and ceramics compared with metallic alloys. There is no fundamental reason why sintered metal should not have comparable mechanical strength to metals, and more will probably be heard of these methods in the future, perhaps as a means of cooling local hot spots rather than of cooling large areas.

A gas turbine combustion chamber with louvred walls through which cooling air is injected has been successfully used in the Auris gas turbine.

Radiation from Luminous Furnace Flames

The luminous flame, from which heat is radiated by small carbon particles produced by heating in a local deficiency of oxygen, is widely used in boilers and industrial furnaces, especially open-hearth steel melting furnaces. The calculation of the heat radiated from such flames is extremely difficult because we do not know the number and size of the carbon particles at each position along the flame. The design of furnaces is often extremely critical, and a better control of flame radiation might well save a lot of money both in fuel consumption and in initial and maintenance cost. In the open-hearth steel melting furnace particularly, the refractory roof has to be renewed frequently, and anything which can be done to lengthen its life while maintaining or even improving the time of melting is of considerable economic value. Since radiation processes do not scale down directly, measurements must be made on flames of comparable size to actual furnace flames. The work of the International Flame Research Organisation at Ijmuiden in Holland, has done much towards the solution of these problems by systematic measurements of radiating powers, temperatures, emissivities, etc., of flames from a variety of burners, conditions of air preheat, etc., with gaseous, liquid and solid fuels. The results so far indicate that for oil fuel diffusion flames, that is oil flames in which the majority of the combustion air is entrained by the flame jet from its surroundings, the flame length is mainly determined by aerodynamic variables and particularly by the momentum of the inflowing jet, and may be studied by model techniques. The radiation distribution over the flame length is dependent mainly on the carbon-hydrogen ratio of the fuel and the initial rate of mixing of air and fuel at the burner. Near the burner the emissivity is high, often approaching unity, decreasing as the carbon burns away along the flame, but since the temperature tends to rise along the flame the actual radiation does not vary greatly along the flame length. Several results of practical value have emerged from these investigations, including the importance of the momentum of the burner jet and the effect of adding pitch creosote to flames of low radiation such as coke oven gas flames, in order to increase their radiating power.

The flame problem is difficult to study because the radiation must be related back to variables such as the burner design and arrangement which are difficult to

specify, the processes of combustion and heat transfer are mixed up in the flame. The only scientific method of approach to such a problem is to make systematic measurements to elucidate the mechanisms at work and so to find which variables have large effects and which have negligible effects.

In gas turbine combustion chambers, flame radiation has only a nuisance value, in that it necessitates more attention to cooling the walls by cold air or other means. Some recent work has shown that the type of flame stabilisers used in gas turbines give flames which radiate differently from diffusion flames. It is probable that the radiation from a luminous flame, unlike that from a non-luminous flame, is capable of control over wide limits by changing the burner design and arrangement even when the fuel and air remain the same.

Cooling of Turbo-Alternators

The cooling of the windings of a turbo-alternator is of great importance especially in view of the uprating of output of a given size which has been demanded in recent years. Conventional methods of air cooling have become inadequate and improved methods using hydrogen instead of air gave better results. Recently, direct cooling of both stator and rotor coils using hydrogen at pressures up to 30 p.s.i. has been developed, in some cases the coolant passing through the conductors themselves which are of tubular form. The reduction in material weight per kilowatt with the consequential increase in maximum rating has been sensational, especially with direct conductor cooling and using high-pressure hydrogen. These improvements have opened up a new range of unit sizes from 200 MW up to 350 MW. It may be said that by far the biggest advance in turbo-alternator design during recent years has come from improved heat transfer.

Liquid Metals

The use of liquid metals for heat transfer has been brought to the forefront by nuclear reactor developments, since their high thermal conductivity makes them suitable for applications where very high heat transfer rates are required as in the fast neutron reactor. Moreover, in sodium and sodium-potassium eutectic alloy particularly, we have substances which do not capture neutrons wastefully and are therefore acceptable for reactors. The use of these substances, especially where water is likely to be present seems at first sight to present formidable practical difficulties, but with the improvements in sealing and handling techniques which are in any case essential in dealing with nuclear appliances, these are no longer considered even serious. Prior to the nuclear application the only liquid metal used was mercury which has the disadvantage of toxicity. Heat transfer coefficients for liquid metals have now been extensively measured, but there is still some uncertainty in interpreting results which may be connected with the effects of thermal conduction in the direction of flow, which is negligible in non-metallic liquids. Liquid metals will probably be more widely used for heat transference in many industrial applications in the future, such for example as in steam reheating in power plant where they would enable the steam to be heated in heat exchangers close to the turbines. It is to be remembered, however, that liquid metals have relatively low specific heats compared to other liquids, which means that the weight of fluid necessary for a given heat removal may be high. The high temperature-drops across the thickness of the walls of the tubes carrying liquid metal coolant may also become a critical factor, reaching as much as 500° F. in 0·02 in. thickness of stainless steel tube wall, for a flux of 3 million B.T.U. per square foot per hour.

In this address I have mentioned only a few of the many kinds of heat transfer problems encountered by engineers and there has been no time to speak of many others such as boiling liquids, rotating bodies, and surfaces in contact. It seems that with the coming of new industries such as electronic, aircraft engineering,

nuclear and chemical engineering, quite new kinds of heat transfer problems arise, or at any rate unexpected developments of old problems, with the result that the scientists and engineers are trying to solve new problems and consolidate the background of the subject at the same time. However, an immense amount of theoretical and experimental effort is now being put into heat transfer research today, especially in the U.S.A., and the recent words of H. L. Dryden, Director of the National Advisory Committee for Aeronautics, may well be quoted:

> Creative research, inventive development, inspired design and novel construction in the field of heat transfer—all must be accelerated on a broad front to keep pace with the needs of modern technology. Nuclear reactors, rocket engines, high speed ballistic missiles reentering the atmosphere, and certain types of chemical engineering equipment encounter extreme conditions of temperature and rate of heat transfer, challenging scientists and engineers to explore new problems and ranges of parameters.

Theoretical methods have greatly contributed to advancing our knowledge in the last quarter of a century, and will continue to do so, but I should like to conclude by making a plea for more experimental work. This can be of two kinds. First, we need careful experiments under controlled conditions, to evaluate physical properties and the experimental constants of phenomena such as turbulent convection. Secondly, we need more experimental observation of what is happening in actual appliances. In heat transfer experimental measurements are not easy, and there are plenty of opportunities awaiting the young research physicist and engineer.

THE STUDY OF FOLK LIFE: AND ITS PART IN THE DEFENCE OF CIVILISATION[1]

ADDRESS BY DR. IORWERTH C. PEATE

PRESIDENT OF SECTION H

I AM deeply sensible of the honour done to me in inviting me to fill this chair. As far as I can discover, only one Welsh-speaking Welshman has previously occupied this position: he was Sir John Rhŷs. Rhŷs held that 'the difficulties of the position of the president of this Section arise in a great measure from the vastness of the field of research which the Science of Man covers. He is therefore constrained to limit his attention, as a rule, to some small corner of it.'[2] This is certainly what I must do today.

'The anthropologist', stated A. R. Wallace in 1866, 'must ever bear in mind that, as the object of his study is man, nothing pertaining to or characteristic of man can be unworthy of his attention.' I am in complete agreement with Dr. D. B. Harden when he stated, from this chair in 1955, that the change of the title of this Section in 1947 from 'Anthropology' to 'Anthropology and Archaeology' 'seems not only unnecessary but misguided. Surely', stated Dr. Harden (and I emphasise his words with complete agreement), 'anthropology—the study of the human race—is a large enough umbrella for us all, whether we be archaeologists or folklorists or students of physical or social anthropology, to shelter under.'[3] Indeed I trust that the present title will never be so interpreted that in our meetings the anthropologist, physical and social, the ethnologist, the student of folk life and the folklorist together will have to share half only of the programme with their colleague the archaeologist, who like them studies only 'some small corner' of the Science of Man.

The 'small corner' which I have chosen for my theme is The Study of Folk Life. We in Wales take particular pride in the contribution made in our country to this study. I hope you will permit me to digress for a brief moment to offer a comment on the reasons for that contribution. In the Welsh tradition, the language occupies a unique position. Though we are often described as a nation of singers, it would be truer to describe us as a nation of writers and of readers, for the arts of poetry and of prose have been given pride of place in Wales for many centuries (to the detriment indeed of the visual arts which generally flourish in urban centres). By the nineteenth century, the Welsh people had become a nation of writers and readers to a higher degree than almost any nation in Europe and in the 1870's and 1880's even book publishers in Glasgow and Edinburgh found it profitable to publish expensive works in Welsh which were sold in their thousands in the Welsh countryside. The Welsh therefore, for generations, have been deeply conscious of their cultural tradition. They have written about it and read about it, in their own language, and have discussed their problems regularly in their village parliaments which were the smithy, the shoemaker's shop, and the carpenter's bench. It was not surprising therefore that at the end of the nineteenth, and in the early years of the present century, the collecting of Welsh bygones (as they were then known) was begun by two or three Welshmen associated with the Cardiff municipal

[1] Address delivered at the Glasgow Meeting of the British Association on September 1, 1958.
[2] *Report of the British Association for the Advancement of Science, Bradford 1900*, p. 884.
[3] *The Advancement of Science*, 1955, **XII**, No. 46, p. 226.

museum. From these small beginnings grew the present national folk collection now housed at St. Fagans.

Before returning to my main theme it is my pleasant duty, too, to pay homage to a former President of this Section, my teacher, Professor H. J. Fleure, who during his tenure of the chair of Geography and Anthropology at the University College of Wales, Aberystwyth, and indeed for many years before the establishment of that chair, directed the attention of many of his students to the study of man and society in Wales and who contributed more than any other individual to the success of the effort to bring together in Wales the various research activities of many workers into an integrated study of Welsh folk life.

You will realise (and what I have said of Wales is proof of it) that the study of folk life is a new discipline and one so far completely unrecognised by all universities in Britain. In making this statement, I do not overlook the existence of the School of Scottish Studies of the University of Edinburgh whose research activities are beyond all praise, or of the excellent Museum of English Rural Life in the University of Reading. But, as far as I am aware, there is no chair of Folk Life or a lectureship in that subject in any British university. In this respect, the subject is in much the same condition in which archaeology found itself less than fifty years ago. So far folk life has been the concern mainly of museum workers, with their emphasis heavily on the material objects only of culture. This has been inevitable, for the many collections of 'bygones' in local, provincial and national museums, looked upon with a certain amount of aloof disdain by some professional museum workers, have been the foundation upon which folk studies have developed.

Let us, in the first place, settle the problem of this word *folk*. It is a much abused word, so abused that some scholars avoid it. Our friends the folklorists are partly to blame for this. When Langland in *Piers Plowman* wrote of a 'fair field full of folk', he had no class distinction in mind, and *gwerin* in Welsh represents the people in general, not a lower class in particular. When W. J. Thoms, in 1846, invented the term *folklore*, it was obvious that he had in mind the lore associated with the hewers of wood and the drawers of water: it was the study of the beliefs, legends, and customs, current among the *common people*; and *folk-song* and *folk-dance* were also so interpreted. It may be that because I come from a people who lay no emphasis on class distinction, I find this connotation somewhat ludicrous. Folk life is the life of the complete community, and folk culture the culture of the whole nation. 'What then is the kernel of the matter?' asks Dr. Sigurd Erixon,[1] 'it is man himself . . . [and] "folk" not man, due to the conviction that it is only by the study of human groups or by comparison between the individual and the group that this science can reach its goal'. Quite simply then, the student of folk life is concerned with the study of every stratum in the community: he is concerned with a branch of general ethnology.

It appears clear, however, that there must be limits to the size and nature of the folk, the community, which we study. George Foster [2] has said that 'one of the most obvious distinctions between truly primitive societies and folk societies is that the latter, over hundreds of years, have had constant contact with the centres of intellectual thought and development'. This is certainly true, as (I think) is his assumption (p. 171) that 'folk cultures will disappear in those places where a high degree of industrialisation develops'. I hope to return to this problem later in this address.

I cannot, however, accept Robert Redfield's definition of a folk society as 'isolated', and 'homogeneous in race and custom'.[3] This conception of homogeneous races is now surely outmoded. Nor can I accept as universally true of folk

[1] 'Regional European Ethnology', in *Folkliv*, 1937, p. 90.
[2] 'What is Folk Culture?', in *The American Anthropologist*, LV, No. 2, Part I, 1953, p. 169.
[3] Ibid., p. 160.

societies Redfield's remarkable description of 'the fabric of society . . . woven of threads of consanguineous and connubial connection'.[1] What is meant here, in plain English, is, I assume, a society based on blood-relationship and marriage. A folk society is not confined to such a relationship. Indeed some of the American anthropologists cloud the issue by differentiating between folk and urban, and folk and peasant, communities. When such special meanings are given arbitrarily to the term *folk*, the exposition of our study is reduced to incoherence.

The study of folk life, as I see it, is the study of the way of life of communities and of nations which are comparatively unaffected by a high degree of industrialisation. The unit so studied can be a nation (as for example Sweden, Ireland or Wales) or it may consist of groups within a national state (as in Germany or the United States). Even in small national units such as Wales or Norway, certain highly industrialised areas and the heavy industries associated with them have to be excluded in part. But there is no exclusion of any class within the community.

> 'Folk' [states one writer][2] means 'all human beings'. You are a member of the folk, so am I. 'The folk' never meant 'yokels'; these . . . were the 'lewde men' as opposed to 'lettred men'. There is ample evidence that what was sung and danced and acted in Langland's day was the same for all, and that in these occupations, castle and cottage met on equal footing: and though dame of high degree made one thing of them and serving-maid quite another, yet both shared one fancy and moved to one rhythm, as their modern counterparts may. What has happened is that song and dance that all knew and practised have survived chiefly among the unlettered who think little of mushroom growths; as have also the customs and courtesies and plain homely virtues of which they are the expression.

Though this quotation refers specifically to folk-song and folk-dance it is universally true of folk life. Let it be said at once that we are concerned with the complete way of life of the community which we study.

Having defined our field as the study of complete communities excluding only the effects of a high degree of industrialisation, we have to determine within what chronological limits we have to work. I attempted a tentative answer to this problem in the *Museums Journal* seventeen years ago.[3] But the problem must be considered afresh. Due to the peculiar way in which folk studies in these islands developed in museum departments of archaeology which also contained collections of bygones, a tendency grew to look upon folk life not as a field of study in its own right but as an appendage to archaeology. And since in practice the archaeologist is generally concerned with the story of man from early prehistoric to medieval times, many persons have concluded that folk-life study is concerned only with the post-medieval period. In several museums this belief still prevails. It has always been a mystery to me why an eighteenth-century wassail bowl should be called a 'bygone' though a Bronze-Age pot is not looked upon as such. The out-dated museum term 'bygones' apparently referred to material not more than three hundred years old and it has been too often assumed that the study of folk life is limited to that same period.

We are concerned with man's mental, spiritual and material struggle towards civilisation and there is no post-medieval or other date-line at which we begin our study. The elucidation of that struggle in prehistoric times depends entirely upon the archaeologist's spade, but his discoveries are of the greatest importance to the study of folk life, and moreover we are of prime importance to him in the interpretation of his finds. Without the aid of the ethnologist any interpretation of the material so excavated would be incomplete. 'Without anthropology', stated Dr. MacIver from this chair in 1932, 'archaeology would be blind of one eye and very

[1] Introduction to H. Miner: *St. Denis: a French-Canadian Parish*, Chicago, 1939.
[2] A. H. Fox-Strangways in *The Observer*, June 19, 1938.
[3] 'The Place of Folk Culture in the Museum', in *Museums Journal*, 1941, pp. 45–50.

short-sighted of the other.' 'It is only through the known that we can apprehend the unknown, only by a study of the present that we can understand the past; and archaeological investigations therefore must be largely barren if pursued in isolation and independent of ethnology.'[1] There must therefore be close co-operation between the student of folk life and the archaeologist, for we begin with the dawn of history and continue with man's story to recent times. Folk life is not an appendage to archaeology: it is a different discipline calling for different methods of approach.

Man's story to recent times: how recent? This is the problem of the date-line at the other end of the story. It has often been said that the student of folk life is not concerned with the period after the Industrial Revolution. But when did the Industrial Revolution begin? Indeed, has it ended? We cannot depend on such inexact terms. When, almost thirty years ago, my *Guide to Welsh Bygones* was published, it had as a subtitle (for which I was not responsible) this phrase: 'A descriptive account of old-fashioned life in Wales.' But when does a fashion become old? Is life old-fashioned in the heart of the countryside and fashionable in the city? It depends upon our standards: from our point of view such a distinction between town and country is quite unscientific. I have already indicated that we deal with communities comparatively unaffected by a high degree of industrialisation, but you will agree that no European community is totally unaffected, in part or in whole, by industrialisation. In my own country of Wales, parts of the south and of the north-east were profoundly affected by industry during the nineteenth century. But even in the most affected parts, evidence of the persistence of the traditional way of life has remained to this day. Mushroom urban growths, like Cardiff, which developed from a small market town into a great cosmopolitan city in a period of less than a hundred years, still show many traces of an old cultural pattern. We have to concern ourselves with that pattern, where it is strong and where it is weak, down to our own day. What we are not concerned with is that overlay of 'sameness' which was borne over a large part of Europe by the great tide of an industrialism which ignored both national and community boundaries. Here again there is no date-line, for despite the weft of industrialism, there are still strands in the traditional warp reaching down to our own day. It is our business to disentangle them. If I may carry the metaphor a stage further, I would say that in our European communities there has been no period of catastrophic change in the history of society where no thrums have been left in the loom to tie the new pattern to the old threads.

It is no part of my purpose today to weary you with a full analysis of the details of our subject. But there is one major consideration which deserves our attention. It is this: what is the character of the material upon which we work? I have already described our field as a study of the way of life of the community, of man's mental, spiritual, and material struggle towards civilisation. Those of us in Great Britain and Ireland who have been concerned with folk studies during the past thirty years have found in making our own classifications of the subject that they correspond closely with those worked out in Sweden and in the other Scandinavian countries. Our debt to Scandinavia is, of course, enormous. It is Sweden and her neighbours who have initiated and developed the study of folk life, and the research work carried out in the three centres—Uppsala, Stockholm and Lund—is the admiration of the whole world. Here I must mention three names, associated in all our minds with those three centres, realising at the same time that several other scholars in those cities have richly bestowed upon us an impressive number of works of scholarship. Uppsala is linked in our minds with the name of the beloved

[1] R. B. Dixon: 'Some Aspects of North American Archaeology', in *The American Anthropologist*, 1913.

Åke Campbell, whose untimely death in 1957 was a grievous blow to folk-life studies in Scandinavia and throughout Europe. His contribution to the study of Lappish problems, to the history of bread, to the study of Irish houses, to cultural problems in Sweden and indeed to a wide variety of folk problems, from 1910 to 1957, is a remarkable achievement. Gay and friendly, he was at all times eager to help promote folk-life studies in Ireland, Scotland, England and Wales; the present status of the subject in Great Britain owes much to his wise counsel and eager enthusiasm. The work carried out by his colleague Professor Strömbäck and himself at Uppsala has established a standard by which most of our labours must be judged. Lund brings to mind the distinguished name of the late Wilhelm Von Sydow, possibly the most outstanding folklorist of the twentieth century. His link with the Keltic peoples of these islands is fittingly remembered by the fact that his magnificent library of folklore is now in the possession of the Irish Folklore Commission in Dublin. To us, the name of Stockholm is almost indistinguishable from that of Dr. Sigurd Erixon, whose seventieth birthday this year gives an opportunity to us all to honour the doyen of folk-life studies in Europe.

One of Dr. Erixon's principal contributions to folk-life research is his study of the Swedish house. The study of houses is basic to our consideration of material culture. In Britain we lag far behind in this direction. Almost every major work so far published has been concerned with castle and mansion to the exclusion of farmhouse and cottage. Nathaniel Lloyd's *History of the English House*, Geoffrey Webb's *Architecture in Britain in the Middle Ages*, and even John Summerson's *Architecture in Britain from 1530 to 1830* are of very little use to the student of folk life. Webb, indeed, dismisses the whole problem of cruck construction in half a paragraph, and that in his chapter on the parish church, while Summerson (despite the title of his book) relegates Scotland to an appendix and ignores Wales altogether! Regional studies in this field are urgently needed. Here may be mentioned the work of Sir Cyril Fox and Lord Raglan—both former presidents of this Section—on houses in Monmouthshire, and the excellent survey work carried out in the north of England by the staff and students of the Manchester University School of Architecture who have set an example which other Schools of Architecture would do well to follow. What is urgently needed at the present time is as full a record as is possible, with scale drawings, of cottages and farmhouses down to about 1900. The interpretation of the material can follow later. I feel that a warning is needed here against a too facile attempt at interpreting the material so collected. The archaeologist, if I may say so, delights in distribution maps, and many geographers, too, have caught the fever. Distribution maps can be dangerous instruments when applied to the study of folk life or indeed always when they are based on incomplete data. A distribution map of, let us say, timber-framed cruck-constructed houses proves nothing except the existence of such houses in the areas shown when the map was made.

There are many other problems relating to houses and house-sites which await elucidation. Take for example the problem of transhumance. Much has been written about the *hendre* and *hafod*—the winter and summer dwellings—in Wales, and comparisons made with transhumance practices in Scandinavia, for example. I suspect that such comparisons are useless and that the migration of flocks and herds in the uplands of Britain must have been a problem of a different character. For instance, you would expect the *hendre*—the winter home—to be always near the valley floor while the *hafod* would be high up. In fact, we find many instances of the *hendre* high on the mountain slopes and of the *hafod* well below. What is first required is to collect the facts. It is obvious, however, that the extreme climatic conditions which made transhumance essential in Norway never existed in our British uplands.

Another problem relating to the history of material culture which needs close

examination is the problem of fashion, in furniture, in dress and in other elements of culture. It has long been the custom to assume that, throughout history, the further a locality may be from London, the greater the time-lag in fashion. We are told that pieces of furniture in the Welsh countryside, for example, exhibiting the same details of craftsmanship as similar pieces in London, are half a century, or even a century, later in date. I cannot believe that, as a general rule, this was at any time true, despite the elaborate theories built up to support the argument. The fifteenth-century literature of Wales tells a story of

> chieftains whose tables were loaded with the choicest of foreign fruits, currants, cinnamon and oranges from the south, and the wines of Rochelle, Bordeaux and Gascony, whose walls were hung with the rich tapestry of Arras, and whose dwellings resounded with the music of harps; the splendour of the monasteries, the 'gold adorned choir', the crystal windows, the lofty roofs resplendent with the bearings of princes, the light of torches and the burning of incense, the rich tombs with sculptured figures and arms of the dead. . . .[1]

Such descriptions, throughout the centuries, do not betoken a countryside lagging behind in fashion. It is always dangerous to postulate a metropolis always up to date in fashion and to assume that the further you travel from it the greater the time-lag in these matters.

I want to interpolate here a brief reference to costume, and to national costume in particular. England, with its national status secure for many centuries, wisely has never boasted of a national costume: its Keltic neighbours have been less fortunate. Far be it from me, here in Glasgow of all places, to venture one word about either tartan or kilt. But you may have noticed frequent references, in recent years in particular, to a national Welsh costume. This costume is always a female costume—the tall chimney hat and the somewhat dowdy flannel bedgown and flannel striped petticoat. It has now been adopted even by modern folk-dancing teams, though I cannot conceive of a less appropriate dress for dancing. None of the elements of this dress is peculiarly Welsh—indeed they have all been found throughout Europe—but they were brought together as a costume through the intervention of a notable patron of Welsh culture, misguided though she was at times, Lady Llanover, who, in 1834, wrote an essay on the national costumes of Wales,[2] in which she urged that the native 'wool packs' be not exchanged for 'bags of cotton' and that woollen garments, the 'native produce', were much better than 'straw bonnets, ribbons, frills, capes, ringlets and all the caprices of fashion'. How much better, she wrote, 'the substantial produce of the Cambrian loom' than 'the thin and comfortless materials' then fashionable. Thus, almost accidentally, a country fashion of the late eighteenth century was put on its way to acquire national status, though the attire of the mere male was completely ignored. Such accidents of history have to be taken into account in any attempt at studying cultural development.

These are some of the many problems which concern us in considering material culture. As far as this part of our study is concerned, I would sum up by suggesting that we should interest ourselves in all aspects of domestic, cultural, civil and craft life; in short, in the organisation of the home, and of the whole community of which the home forms a part.

So far I have spoken only of 'the visual material witnesses of the long, slow imperfect growth of mankind'.[3] But houses and furniture, ploughs and carts are

[1] W. Garmon Jones: 'Welsh Nationalism and Henry Tudor', in *Trans. Hon. Soc. Cymmrodorion*, 1917–18, p. 2.

[2] Gwenynen Gwent: *The preservation of the Welsh language and the National Costumes of Wales*, London, 1836.

[3] A. M. Buck: *Presidential Address to the North-western Federation of Museums and Art Galleries*, 1958, p. 15.

only mute witnesses to man's development. They tell us a little, but only a little, about their users. What of the names men have given in different places, at different times, to the furniture and utensils of the home, to the implements with which they tilled their fields, and to the knife and rasp of the craftsman? What songs did the ploughman sing? What of men's feast days and days of sorrow? What thoughts had they about the universe and an after-life and how were they expressed? Indeed, what of their whole language? My old friend, the late Professor W. J. Gruffydd,[1] once said that

> dialect is the biological source of all living speech, the only medium through which such speech can absorb the necessary vitamins without which it will languish and die. For ... a 'dead' language is not necessarily a language like Latin or Hebrew, that is no longer spoken. A language may be on the lips of millions and yet, to all purposes other than the lowest, be as dead as a doornail; which is indeed a very apt comparison, since the nail moves with the door, but has no motion of its own.

He went on to deplore the disastrous effect of the terrible Americanisms of the talkies upon spoken English, and added:

> It is disturbing to think that the council school has only levelled the beautiful old native dialects into a uniform flatness in order to prepare the way for raising upon it this unlovely edifice of American vulgarism. . . . Unless we are careful, we are in danger of losing the English language as a medium of inspired expression. That is to say, it may become for England, and a great part of the world, what Latin had become for mediæval Christendom—a nerveless, dull, undistinguished *lingua franca*, adequate for the scholar and the scientist, but impossible for the poet and the prophet. If our emotions, as expressed in language, have only drawing-room associations, if they have no associations with the field and the factory, with the country church and the little township, it will be impossible to clothe them in anything that might be described as exalted language. It is time that we should revise our notions of ignorance and illiteracy. We have a heritage to preserve. . . . .

It is for these reasons that a full survey of all the dialects of these islands—English, Welsh, Lowland Scottish and Gaelic—is urgently needed. One of its results may be to revitalise all our languages.

All these problems concern the student of folk life. The study of material culture is part only of our task: we have to deal with the complete life of man and the community. Too great an emphasis cannot be placed on this urgent need for studying dialects and recording them. This includes not only an examination of local dialects, their characteristics and their boundaries but, through such an examination, the compilation of a whole host of vocabularies—the vocabulary of the hearth and home, that relating to domestic animals, from the cat and dog to the horse and cow, the vocabulary of agriculture in all its stages, of the festival and the market, of church and chapel, of birth, courtship, marriage, and death. The twentieth century has seen cataclysmic changes in these directions. The horse-drawn plough has given way to the tractor plough and if the nomenclature of the old traditional implement is not recorded now, a great part of the wisdom of the ages will be irretrievably lost. The intrusion of radio and television into the remotest parts of these islands is itself causing a swift and silent revolution. In England the dialects which have provided the literary language with its lifeblood are being standardised if not emasculated: in Gaelic-speaking Scotland and in Welsh Wales, the languages are being suffocated.

It appears to me that all the possible resources of the state should be brought to support the work on oral traditions carried out in Scotland by the School of Scottish Studies in the University of Edinburgh; in England by the University of Leeds; and by the Welsh Folk Museum in Wales. In Ireland, the magnificent work of the Irish Folklore Commission is subsidised by the State. But what do we

[1] W. J. Gruffydd: 'Yeoman's English'. Broadcast talk, March 22, 1935.

find in Britain? In Scotland the plea of the School of Scottish Studies for an emergency grant to enable steps to be taken immediately to record and examine the traditional culture of Benbecula and the Uists (in view of the government's decision to develop a guided missiles range there) was summarily rejected by the Secretary of State. This 'irresponsible attitude' (the words are those of Professor Oftedal of Oslo[1]) may deprive the scholar of a wealth of material of untold importance to the history of civilisation in these islands. In Wales, where large tracts of land have been compulsorily acquired during the past twenty years and their cultural patterns completely destroyed, we are struggling desperately to collect the funds necessary to record the language and traditions of the area left to us. It is a sobering thought that the price of one guided missile would more than solve all our problems in this direction.

Here I come to the second part of the title of this address. What is the rôle of the study of folk life in the defence of civilisation? It is no part of my duty, in this chair, to be concerned with politics. But in an Association for the Advancement of Science we *are* concerned in a vital way with civilisation, for it is civilisation, bringing man out of barbarism to a higher level of life, which is the basis of our study. In this twentieth century our so-called civilisations are a nightmare. They have no longer the grandeur of the civilisation of an earlier age. The problem troubled A.E., that great Irish thinker, forty years ago. 'Our civilisations', he wrote,[2] 'grow meaner and meaner as they grow more urbanised.' He describes city life:

> the spirit grows heavy as if death lay on it while it moves amid such things. And outside these places the clouds are flying overhead, snowy and spiritual as of old, the sun is shining, the winds are blowing, the fields are green, the forests are murmuring leaf to leaf, but the magic that God made is unknown to these poor folk. The creation of a rural civilisation is the greatest need of our time. . . . The fight now is not to bring people back to the land, but to keep those who are on the land contented, happy, and prosperous. . . . If we read the history of civilisation we will find people in every land forming little clans co-operating together. Then the ambition of rulers or warriors breaks them up; the greed of powerful men puts an end to them. But, whether broken or not, the moment the rural dweller is left to himself he begins again, with nature prompting him, to form little clans—or nations rather—with his fellows, and it is there life has been happiest.

It is the story of 'this eternal task of building up a civilisation in nature' that we study, in England, in Scotland, in Wales, in Ireland, in Norway and Sweden, in Holland, in all the 'little nations' of Europe. The sum-total of knowledge and of experience garnered in and by those communities is the greatest riches which this world will ever possess. To wipe out the accumulation of experience which is, for example, the culture of the Gaelic-speaking Hebrides, and moreover to allow it to be done without any record of it, would be a return to barbarism, a reversal of the true process of civilisation.

Henry Adams[3] was right when he observed that 'bombs educate vigorously, and even wireless telegraphy or airships might require the reconstruction of society'. In the brief forty years since he wrote, the disintegration of western society has been accelerated. We live in the era of the war machine, the finance machine, the industrial machine and the education machine, in the era of technique for its own sake, a life 'sweet in the mouth but bitter in the belly'. Lewis Mumford has examined the underlying barbarism which is our present condition.

> In the passive barbarism that the United States now boasts under the cover of technical progress, there is no promise whatever of victory or even bare survival. Without a deep regeneration and renewal the external triumph of American machinery and arms will hasten the downfall of the Western World.[4]

[1] *Gwerin*, I, 1957, p. 147.
[2] A. E.: *The National Being*, Dublin, 1920, pp. 63–4.
[3] Henry Adams: *The Degradation of the Democratic Dogma*, New York, 1919.
[4] Lewis Mumford: *The Condition of Man*, London, 1944, p. 376.

Over against the passive barbarism of the West we have in eastern Europe the active barbarism of a politico-religious movement which seizes power and holds it at all costs, 'no lies, brutalities, frauds, defamations or self-contradictions standing in the way of this aim'.[1] These are the two forces threatening to reduce the world, if not to complete destruction, at least to the level of brute sensation.

Between these two colossal world forces, with 'their over-charges of empty stimuli, their perpetual miscarriage of technique, their materialistic repletion, their costly ritual of conspicuous waste, their highly organised purposelessness', [2] are the 'little clans', the 'little nations', those communities, principally rural, which from times ancient beyond memory, have accumulated traditional cultures, mellowed and ripened by the wisdom of countless generations. Their tradition is the only buffer against the new barbarism. This David must conquer the new Goliath: as that great Scotsman, Patrick Geddes, used to say: 'Animals are always attempting the impossible and achieving it', and that, too, must be our essential condition. It is the life of these small communities which we study and record throughout Europe. Their traditions are the field of folk-life research: it is their cultures that we must save. For, whenever a little community is submerged by the new barbarism, a bastion of civilisation falls. If the fall is inevitable (and that I do not concede) at least we must not fail in preparing a full record of that which perishes. That is our duty, and by carrying it out faithfully and well, we play our part in the defence of civilisation.

It seems to me therefore that the study of folk life has a serious meaning and a real purpose in our modern world, a purpose beyond the mere limits of an academic discipline. I trust the day is not far distant when throughout the ancient world there will be centres for the study of folk life. Consider the wealth of material still unrecorded throughout the many countries of Europe and the Near East, the local cultures of Christian nations and of Islam. All records of these should be preserved in archives at recognised centres. Their accumulation at such centres of research would enrich scholarship and true civilisation for centuries to come.

Let me sum up. The present 'upthrust into barbarism' (which is a fair definition of the machine age) represents the disintegration of civilisation, the abandonment of the countryside and the growth of huge suburban wens. But neither the machine on which our urban growths are founded nor automation creates tradition. What is transmitted from generation to generation is the many basic skills of individual men. And it is with skills, with manual operations, with spiritual beliefs, in the pre-megalopolitan cultures, that we are concerned. Our present mechanical civilisation is not an absolute. As has been well said, 'all its mechanisms are dependent upon human aims and desires'.[3] Those of us who study folk life can contribute not only to the sum of scientific knowledge but can also assist in the rehabilitation of a society degraded by the arrogance of the techniques which it has devised. It is my belief that if our civilisation is to be saved, the study of the 'little clan', its social organisation, its skills, its spiritual beliefs and its visions, will play a vital part in that salvation. As students of folk life, it is by that faith that we should work.

[1] Mumford: op. cit., p. 341.
[2] Ibid., p. 380.
[3] Lewis Mumford: *Technics and Civilization*, London 1946, p. 426

SOME ASPECTS OF CELLULAR DIFFERENTIATION[1]

ADDRESS BY DR. E. N. WILLMER

PRESIDENT OF SECTION I

THE problems of cell differentiation are as wide as the biological science themselves, so it is evident that my address can only attempt to discuss an insignificant proportion of the whole problem. Nevertheless I believe that the time has come when by some reorientation of our ideas and by attempting certain not immediately obvious correlations it may be possible to suggest problems for research in directions which have not hitherto been followed, or at any rate not extensively so, and which may perhaps in the end lead to a better understanding of the events which occur at the cellular level during the development of an animal. My purpose in this address is not therefore to review a field of accomplished work, but to try to present certain somewhat different and possibly new approaches towards unknown territories whereby a better understanding may be obtained of some of the events which cause the tissues and cells in the body to behave in the way in which they do.

My story starts with collagen fibres. Most medical students at an early stage in their career associate these tough white fibres with the loose connective tissues or fascia of the human body, and thereafter they lose all interest in them, apart from using them in their tanned form as leather in replacing their footwear from time to time. Most zoologists hardly know of their existence. Nevertheless there are exceptions, and it has now been shown that collagen is found in most animals even in those as primitive as the sponges and coelenterates. In fact, collagen fibres enter into the structure of a great majority of animals and, together with certain mucoproteins and mucopolysaccharides, are the product of a particular form of cell, called the fibroblast. These extremely strong fibres form an obviously useful building material for animal bodies so that it is not surprising that, once they were developed in evolution, they should have persisted. Nevertheless this usefulness does not tell us how they first started. Collagen fibres, as proteins go, are relatively simple in structure and have a 'backbone' in which three amino-acids, glycine, proline and hydroxyproline predominate. They appear to be liberated from or formed by cells in the surrounding fluid as relatively short rods (tropocollagen units) about 3,000 Å long, 14 Å wide and with a molecular weight of about 340,000. These units, when given the right conditions, join together, end to end and side by side, in a special manner to form characteristically cross-banded fibres. The cells which produce tropocollagen must thus be concerned first with forming or eliminating these three amino-acids in appreciable amounts. Not only do fibroblasts produce tropocollagen but it seems probable that such cells can also liberate into their surroundings muco-substances, built up from sugars, amino-sugars, and uronic acids and these are sometimes combined with proteins. Many of these muco-substances (mucoproteins or mucopolysaccharides) are normally combined with sulphate, e.g. the ground substance of cartilage. Like the collagen fibres, these substances are formed extracellularly or are immediately released from the fibroblasts and are not stored intracellularly by them, as they may be by cells of other kinds. When stripped to its skeleton then the question which has to be answered in connection with the formation of the main supporting tissues is why certain cells, i.e. the fibroblasts, habitually form or excrete glycine, proline and hydroxyproline on the one hand, and sugars, amino-sugars and uronic acids

[1] Address delivered at the Glasgow Meeting of the British Association on September 1, 1958.

on the other. Unless they did this, skeletal structures like bone, cartilage and fascia might never have developed. Of these two groups of substances typical cross-banded collagen fibres are almost exclusively produced by fibroblasts, though certain epithelia, e.g. the epidermis of the earthworm, may secrete somewhat similar material with the same high-angle X-ray pattern as collagen, but which differs in low-angle X-ray pattern, and in its appearance under the electron microscope. Other types of cells than fibroblasts may also eliminate polysaccharides, or amino-sugars, and still others store them within their cytoplasm, but, again, fibroblasts habitually produce them and weave them into their exclusive products, e.g. chondrin, ossein, etc. The 'ground substances' of the connective tissues of animals certainly represent a very large extracellular output of muco-substances. Because of their universal occurrence in animals the suggestion is very strong that collagen fibres and organised ground substances, though certainly primitive, are probably not in fact the most primitive cellular products from an evolutionary point of view but rather represent the secondary utilisation of some regularly produced and even more fundamental by-products of cellular activity. The real problem, therefore, in this aspect of the process of differentiation is to discover the nature of the reactions within the cells which give rise to the by-products which have subsequently been so widely utilised for structural purposes. Some or all of these primary by-products may be, and indeed probably are, produced by other types of cells and are not necessarily the exclusive property of fibroblasts, but it is probably only the fibroblasts which have been able to turn them to advantage. Some hint of this may perhaps be obtained by reference once more to the sponges. It so happens that the scleroblasts, which are the fibroblasts of the calcareous sponges, all arise from one group of epithelial cells in the embryo and it is therefore possible that a clue to their nature may be derived from a study of these embryos, and in particular of the cells which are the precursors of the scleroblasts.

At a rather early stage in the development of calcareous sponges the hollow spherical blastula is divided into two distinct portions—an anterior (or animal) portion composed of columnar flagellated cells and a posterior (or vegetal) portion of cubical, granular and phagocytic cells; in some species there may also be an intermediate group of columnar non-flagellated cells containing some granules. The scleroblasts or fibroblasts are believed to develop from the flagellated cells, so that these cells are likely to be the ones whose metabolism can give rise to some at least of the by-products which are subsequently used for connective tissue formation, and morphology alone would perhaps suggest, by analogy with other epithelia which display two types of cell, that the metabolism of the flagellate cells differs rather fundamentally from that of the posterior phagocytic cells.

These blastulae of the calcareous sponges can probably be considered with a fair degree of truth to be similar in character to the most primitive multicellular organisms or earliest metazoa. In whatever way we may consider that the metazoa evolved from the protozoa, the immediate product is likely to have been either a solid 'morula' or a hollow 'blastula' and on account of the necessity for adequate oxygenation of all the constituent cells, etc., the blastula might be considered to have certain advantages, and indeed it is the more commonly found in embryology. It is not always composed of two visibly different classes of cell but, very often there is other evidence that it is divided into an animal pole and a vegetal pole whose metabolic processes differ. For the present discussion particularly, but also on general grounds, the nature of this difference between the two sorts of cell thus becomes a problem of great interest and basic importance. It would certainly be one worth further investigation in those sponge embryos where the two groups of cells can be sharply differentiated from each other.

However, another approach is possible. There exists in a particular amoeba, *Naegleria gruberi*, a single-celled organism which, in a matter of minutes, can

change from a creeping phagocytic amoeba to a free-swimming flagellate. The change is freely reversible, so that *Naegleria* shows from time to time the same sort of difference as occurs from place to place in the surface of the sponge embryos. Moreover, these amoebae have, under certain conditions, the capacity to adhere together, so that in this organism all the requirements for the formation of a simple colonial blastula-like structure appear to be potentially present. This being so, the cause and nature of the amoeboid-flagellate change of phase in these amoebae may be relevant and fundamental to our understanding of the significance of the two types of cell in many of the free-swimming blastulae. In the past, the occurrence of two types of cell in so many blastulae has been thought of as giving advantage to the organism because the flagellate cells could aid the dispersal of the species while the amoeboid and phagocytic cells could provide for greater food storage and a wider range of food materials by allowing the engulfing of solid particles. This may well be so, but the observations on amoebae suggest that other factors may also be important. In the amoeba, the indications are that the ionic content, and particularly the cation content, of the surrounding medium is the determining factor in effecting the metaplasia. Broadly speaking and with a great deal of over-simplification, the amoeboid form appears to be adopted when the external medium is ionically more concentrated than the organism, i.e. in NaCl concentrations higher than about M/20, and the flagellate form when the cell is in danger of losing cations to the surroundings. Moreover, evidence is accumulating to show that the flagellate form has a mechanism, which, incidentally, can involve glutamic acid (a precursor of hydroxyproline and a component of collagen) by which the cell actively accumulates cations from a medium more dilute than itself and so returns to the amoeboid form. These investigations on *Naegleria*, while still very much in their infancy, put quite a new face on the problem of the primitive blastula. They emphasise that ionic balance looms large in the stability of such an organism and that mechanisms for the control of the water and ionic content of the organism are a necessity for its stability.

A hollow epithelial sphere of identical cells is probably not a very stable structure osmotically. Indeed the evidence from certain tissue culture experiments with fragments from the chick's kidney shows the dangers very well. When small pieces of such kidneys are cultivated in a suitable medium many of the cut tubules close themselves off at their cut ends and so become little hollow vesicles. Under those conditions the proximal tubules may be made to swell up while the distal tubules simultaneously may collapse and their lumen disappear. Stability in such a system can be achieved either by a more or less complete impermeability of the wall, a situation which in the long run would not be very suitable for a group of metabolising cells, whose waste products might be harmful, or by regulating the water and salt movements in and out of the sphere. This might be most effectively done by balancing the inpushing action of one type of cell against the extracting action of another. The first situation is probably approached in the early blastula of the echinoderms: in such a blastula the cells are at first all of one type and it can be shown that glucose, for example, is not picked up from the surrounding sea-water till the blastula becomes free-swimming, by which time there is generally a visible differentiation of cells at the vegetal pole. The second device of balancing one system against another is very frequently adopted by nature when delicate regulation is required and many secretory mechanisms seem to work on principles of this kind. Unlike the kidney vesicles in the chick tissue cultures, normal blastulae do not habitually either distend or collapse except at definite times and in an orderly fashion, as in gastrula formation. Blastulae are therefore presumably osmotically and ionically stabilised and changes in this stability may sometimes be the key to movements of gastrulation, etc., either directly by pressure changes caused by fluid accumulation or extraction, or, more probably, since movements of

invagination in the sea-urchin, for example, are not affected by opening the blastula cavity, by alterations in cell form connected with the directional ionic and water movements.

The implication of this argument is therefore that the blastulae of these calcareous sponges are composed of a single layer of epithelial cells which fall into two main groups and that these groups, on the analogy of *Naegleria*, are concerned at least partly with the water and ionic relationships of the organisms as a whole. If these two forms of cell activity are stable and capable of being passed on to the next generation of cells then every cell in the organism, as it develops further, must have initially belonged to one of these two categories and so inherit its characteristics. If, as in *Naegleria*, the change of cell behaviour is reversible, then any cell may be able to enter either state with more or less ease according to its local environment. In the calcareous sponge the establishment of two cell types occurs early in development, and unless artificially disturbed is probably fairly stable. In other species, e.g. sea urchins, such differences between cells are not apparent till later, and, continuing this argument further, it is possible to imagine that some cells may remain labile, if they are in a stable environment, till late in development and only differentiate into one form or the other at a very late stage. At this point we may leave sponges and amoebae for a time, and extend the idea of ionic stability by opposed action to an examination of some epithelia in higher organisms.

The surface of an animal is of vital concern to the animal, in relation to the gain or loss of body fluids. An animal could maintain a passive impermeability by building an impervious wall or it could maintain an active impermeability by regulating intake and output. Our own skin of keratinising epithelium approaches the former type with the possible minor exception of the sweat glands. The skins of the frog and eel approach the latter type. They regulate the water passing through them and in the case of eels and other fish we shall hear this morning that special cells can be developed for the regulation of certain aspects of the ionic content of the animal with reference to its environment. We shall also hear this morning how the passively impermeable keratinising skin of the chick can be converted by high concentrations of vitamin A into a membrane composed of two distinct types of cell, thus perhaps becoming capable of maintaining active impermeability, or of controlling the fluid environment of the cells.

If we now follow up this concept of actively controlled permeability and the implication of the existence of the two sorts of cell which are produced by the vitamin A treatment of chick skin, e.g. ciliated cells and mucus-producing cells, some interesting points emerge. The cells are the same types as those which occur naturally in the epithelium of the trachea, and in the embryonic oesophagus, and again lack of vitamin A in the trachea and age in the oesophagus can alter the epithelium back to the stratified and even to the keratinising type. The trachea is lined by a membrane which may have to deal with greatly varying concentrations of fluid on its surface according to the temperature and humidity of the air which is being breathed. Regarding the membrane then as a barrier between the outside world and the fluid contents of the body the control of the fluid on this surface is important and the problem before the cells of the epithelium is very similar to that of the cells of a blastula, and it will be remembered that there were two sorts of cell in the blastula, i.e. flagellated and 'phagocytic' cells. It is therefore possible that the cells of the tracheal epithelium, ciliated and mucous, represent a similar symbiotic partnership which by the antagonistic actions of the cells of the two classes constitute a fluid- and ion-regulating system, capable of counteracting the effects of excessive hydration or of desiccation which external conditions may impose on the epithelium. On this hypothesis the ciliated cells are thus killing two birds with one stone: they are providing a current of fluid over the surface in order

to keep the membrane clean and they are also helping in regulating the ionic equilibrium. Similarly the goblet cells are adding the slimy water-retaining mucin to the secretion on the one hand and opposing the action of the ciliated epithelium in relation to the ionic equilibria on the other. It is not possible at this stage to be more precise about the methods by which the two types of cells may regulate the ionic concentrations. In the first place there are at least four cell surfaces involved in the situation; the ciliated and goblet cells each present surfaces to the lumen and to the tissue fluids. Secondly the ionic equilibria may involve active movements of water, cations, anions or other dissolved substances and the active movements

TABLE I

Epithelium	*Cells*
Skin	Keratinising, dendritic
Sweat glands	'Dark', 'light'
Mammary gland	Acinar and duct
Salivary glands	Mucous, serous, duct
Stomach	Surface mucous, neck mucous, peptic, oxyntic, argentaffin, argyrophil
Intestine	Goblet, brush-border, Panneth, argentaffin, argyrophil, mucous cells of Brunner's glands
Pancreas	α, β, exocrine, duct
Liver	Duct, hepatic (note also Kupffer cells)
Trachea and bronchi	Ciliated, goblet, basal, non-ciliated
Choroid plexus	Polypoid, ciliated
Kidney	Capsule, neck tubule, proximal tubule (segments 1 and 2), 'thin' loop of Henle, 'thick' loop of Henle, distal tubule, collecting tubule
Fallopian tube	Ciliated, non-ciliated
Uterus	Ciliated, non-ciliated, mucoid
Retina	Rod, cone
Olfactory organ	Hair, supporting
Gustatory organ	Hair, supporting
Auditory organ	Hair, supporting
Anterior pituitary	Basophil, acidophil, chromophobe
Thyroid	Chief, P.A.S.-positive
Parathyroid	Chief, eosinophil

almost inevitably bring passive movements in their train. Finally it may be noted that the basal cells of the epithelium are probably still in the labile state, to which reference has already been made (p. 98), and from which they may develop either into the ciliated or the goblet cell.

If this idea of balanced activity between pairs of epithelial cell types is true for the trachea, it may well be true for other epithelia also, particularly those which separate an external fluid from the tissue fluids of the body. Table I shows the cell organisation of a variety of such epithelia in the mammalian body and it is certainly clear from the table that the majority of epithelia contain at least two demonstrably different types of cell. The nature of these differences is therefore worth further study and in this connection some recent observations with the electron microscope may be important. Many of the complexities of these epithelia

are, of course, associated with the production of special products, but these may be secondary developments and the fundamental properties of the cells which can give rise to the special features are the qualities which require study. For example, the organisation and function of the acid-producing oxyntic cells in the stomach suggest that they are probably very different in origin from the pepsin-producing cells.

As indicated above, the movement of water or ions through an epithelial cell, in contrast to an isolated cell or protozoan, must presumably involve passage through that part of the membrane of the cell which lies adjacent to the surface or lumen and also passage in the opposite direction with respect to the cell surface through the basal membrane next to the tissue fluids. The structure of these surfaces has recently been receiving attention particularly from the electron microscopists, and some of the results may be diagrammatically represented in the series shown in Fig. 1 and such surfaces may be compared with those of free-living cells and protozoa. On the right we have the flagellate cell with an otherwise simple cell membrane; on the left we have a cell with lobose pseudopodia or with gross invaginations of the surface as in amoebae or in the macrophages of mammalian tissues. In different epithelia various gradations between the two extremes are represented on one or other of the effective surfaces. From the figure the distribution of some of these cell types in various epithelia can also be studied and the suggestion could be made that two cells are selected by nature for each type of epithelium and that they are chosen appropriately to the composition of the fluids on either side of the epithelium. The more widely varying the fluids the wider apart in the series are the members of the respective pairs; the more stable the composition, the nearer together can the cells be in the series and the more delicate the balance. The particular cells chosen would depend on the composition of the surface fluid with respect to the local tissue fluid.

Cells which in embryology leave the primary epithelia of the organism to become cells of the connective or other tissues presumably inherit something of the original characters which they possessed in the primary blastula epithelium, and it is tempting to regard fibroblasts, for example, as derivatives of the rather orientated cells on the right of the diagram, i.e. those which tend to have cilia or flagella. The amoeboid macrophages on the other hand would then be considered as derivatives of those surface cells which are more akin to the cells on the left of the diagram, whose cytoplasm appears less fibrous and whose surfaces are more plastic. The action of vitamin A is again of interest here, because it certainly acts on fibroblast-like cells in that excess of it has been shown to be capable of rendering the mucoprotein and cartilaginous matrix of developing bones completely soluble. The vitamin A must therefore do something to the cell's surface or metabolism which affects the manner in which polysaccharides are formed or excreted by the cells. Nor does the story end there, because the particular form of vitamin A is apparently a matter of great concern to animals which live in salt or fresh water. Vitamin A_1 is characteristic of terrestrial and marine animals and their eyes depend for their visual pigment on the corresponding derivative of vitamin A_1, namely, visual purple or rhodopsin, while fresh-water animals have vitamin A_2 with visual violet or porphyropsin in their retinae. These visual pigments are interesting in themselves and may, in certain cases, because of their different spectral sensitivities suit the animals living in the sea or in fresh water, but there is no obvious reason why life in the deep blue sea or on dry land should require a different pigment from that required for life in ponds and rivers. Presumably the type of vitamin A, which affects not only the eye but also the liver and many other tissues, is important for other reasons than vision, and the ionic content of the environment and the necessity for the animal to maintain a proper water balance in its body spring to mind as the features in the situation which should be linked with the vitamin. Here then it

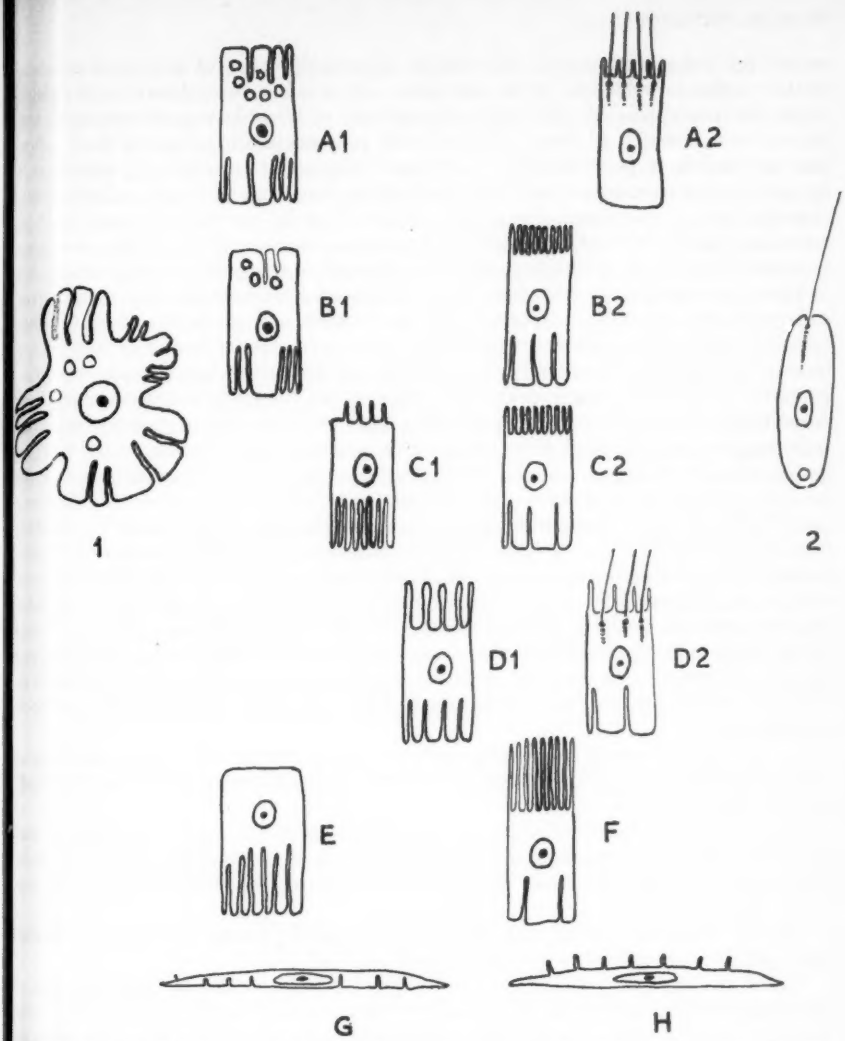


Fig. 1.—Simplified diagrams of electron-microscopic structure of cell membranes, showing the distribution of cilia, microvilli and infolded membranes. The interdigitations with neighbouring cells have been omitted. In the epithelial cells the upper surface is the free or distal surface and the lower surface is the basal or proximal surface adjacent to the tissue fluids. In E the infolded membranes lie next to the aqueous humour from which they are separated by a 'basal membrane.'

1 Amoeba
2 Flagellate
A Trachea
  1 Goblet cell
  2 Ciliated cell
B Intestine
  1 Goblet cell
  2 Brush border cell
C Kidney
  1 Distal tubule cell
  2 Proximal tubule cell
D Choroid plexus
  1 'Polypoid' cell
  2 Ciliated cell
E Ciliary epithelial cell from the eye
F Epididymis cell with 'stereocilia'
G Endothelial cell
H Cell from peritoneal surface

seems likely that vitamin A, and indeed a particular type of it is in some way related to the ionic balance of the organism just as it has been shown to be related to the form of epithelial cells and to the activity of fibroblasts and bone cells and the manner in which all these cells deal with polysaccharides. Does it then follow that movement of polysaccharides and ionic movement through cell membranes are closely interconnected, and that the polysaccharides and tropo-collagen units characteristic of connective tissues have arisen from by-products of those cellular activities which were originally directed towards ionic stability? In the enormous variation of structure throughout the animal kingdom the problems of permeability of the various surfaces which have been developed have undoubtedly been solved independently in each individual case, and there are probably all gradations between completely 'passive' permeability or impermeability and of greater or less 'active' permeability. Passive impermeability may depend on keratin, oils, or muco-proteins, and active permeability may depend on pumping mechanisms whose constitutent parts still need investigation, but which, as the arguments outlined here suggest, may involve polysaccharide precursors which in some way work in conjunction with certain amino-acids. Perhaps the recent observations on the amination of sugars by an enzyme from yeast, working in the presence of glutamine and ATP with the formation of glutamic acid, has some significance in relation to this problem. Perhaps also the dependence of the uptake of potassium by certain bacteria (e.g. *Streptococcus faecalis* and *Staphylococcus aureus*) on the simultaneous uptake of glutamic acid and the fermentation or oxidation of glucose is another pointer; and this again is perhaps linked with the observation that slices of brain tissue cease to lose potassium to their surroundings if provided with glutamine and glucose. Certainly there is a *prima facie* case for associating the muco-substances, collagen precursors, and the movement of ions and water through cells and their membranes.

In summary, I would like to emphasise several points which emerge directly or indirectly from this discussion and which seem to me to be important in the study of cell differentiation.

First, a cell has to preserve not only its osmotic but also its ionic equilibrium, and probably the more specialised a cell becomes, e.g. for nerve conduction, muscular contraction, internal secretion, etc., the more delicately has that ionic equilibrium to be adjusted.

Second, both the internal and external ionic equilibria must therefore be stabilised and they are probably peculiar to each type of cell.

Third, the opposing actions of two or more types of cells on their immediate environment may often be more effective as a controlling mechanism than the action of cells of only a single type. Symbiotic or co-operative pairs and systems should therefore be expected, and there is ample evidence to suggest that they actually occur.

Fourth, a cell in preserving its own ionic equilibrium alters the immediate environment for neighbouring cells,

(*a*) by addition or removal of the ions in question
(*b*) by addition or extraction of metabolites used in the process

and these subsidiary actions may form the basis for 'evoking' a different form of activity from neighbouring cells. Mesenchyme cells eject into the medium surrounding them certain precursors of mucopolysaccharides and fibrous proteins, and it is now well established that mesenchyme cells have definite effects in the induction or evocation of various forms of epithelial behaviour. For example, they cause the epithelial cells of kidney or salivary gland rudiments in tissue culture to begin to differentiate or at least they allow them to display their potentialities. The observations on *Naegleria* show that in this organism the cell may change its form

in direct response to ionic changes in its environment and thus, in the higher animals, such influences may also be at work, either alone, or in combination with other factors. Some 'evocations' may thus be caused by alterations of the ionic equilibrium of the cells in question.

Fifth, the immediate local environment of each cell differs more or less from the local environment of every other cell, and each cell adapts itself to its own special local environment. Differentiation therefore should probably be seen as a process which for any given cell is a very intimate and delicately-balanced cellular phenomenon closely dependent on the immediate surroundings of that cell. A tissue as a whole may be a relatively stable system, but when its cells are isolated each is removed from its own personal niche and must either re-adapt to its new situation or perish.

Sixth, a cell in the animal body seldom has only one local environment; many epithelial cells, for example, have one environment at their basal ends, another at their distal (or luminal) ends and a third provided by the cell surfaces of adjacent cells in the epithelium. Such cells are like sacks of corn in a granary—the mice attack them from the floor, the weevils from the air, and moulds grow between the sacks.

Much of what I have said will inevitably prove to be wrong and in any case it must be regarded as frank speculation. To that extent it is, of course, bad science. Nevertheless the British Association is the place where workers in several disciplines can meet and exchange ideas and I have in this address attempted to introduce some ideas which, though nebulous and unorthodox, may nevertheless tempt biochemists, histologists, embryologists and zoologists to consider the problems of cell form and differentiation on a somewhat wider basis and to co-operate more closely with each other in attempting to solve them.

A REVIEW OF SOME PROBLEMS OF COLOUR VISION AND COLOUR BLINDNESS[1]

ADDRESS BY PROFESSOR R. W. PICKFORD

PRESIDENT OF SECTION J

In November 1907, fifty-one years ago, the University Court of the University of Glasgow agreed to institute a Lectureship in General and Experimental Psychology. Henry J. Watt, a distinguished member of the Wurzburg School of Psychologists, was appointed to the Lectureship from January 1908, and also to take charge of the new Laboratory of Experimental Psychology, then attached to the Physiology Department. The Laboratory was equipped in part by a donation of £450 in 1908, and a gift of auditory apparatus from his collection, by friends of John G. McKendrick, who had been Professor of Physiology in the University from 1876 until 1906 and was much interested in the special senses. In 1909 the Psychology Department (and Laboratory) were in the Faculty of Arts, and this has been our home ever since. The Department has always been associated with work on perception and the special senses, and it is appropriate that this Address should review some of the work carried out here and elsewhere on colour vision and colour blindness, and its applications in occupational psychology, and industry.

The History of Colour-Blindness Tests

The importance of adequate tests for colour blindness became recognised as a result of the widespread use of colour signals on the railways and for shipping. One of the first tests was Holmgren's Wool Test, and it showed about 5 per cent of men to be red-green blind. Other wools tests have been used, such as Loken's Nela Wools Test.[2] Edridge-Green showed that the Holmgren test passed a considerable number of colour-vision defectives, and he devised a coloured beads test, a lantern test and a colour perception spectrometer. He stressed the importance of colour naming in addition to colour matching in tests for colour blindness. His spectrometer had a revolving drum showing the wavelengths of the lights viewed, and a successor to it, perhaps, was Ramsey's spectrometer, which had two independent slits, the lights from which could be viewed one above the other for comparison, or combined for the identification of complementary colours. It had two dials reading wavelengths, and a white light source for desaturation when necessary. Edridge-Green's lantern was designed to meet the objections of certain people who said that their business was with signal lights and not with coloured wool. It is a lantern with a small aperture which can be varied in diameter and adjusted to the size of a signal lamp at a specified distance. The colour of the light spot can be varied through a wide range of hues and saturations by means of glass filters, and also made to appear as if seen through rain, mist or fog by the use of ribbed, frosted or smoked glasses. The beads test is interesting but not very useful.

[1] Address delivered to Section J (Psychology) at the Glasgow Meeting of the British Association on Thursday, August 28, 1958.

[2] Most of the references in this Address will be found in the following publications: *Individual Differences in Colour Vision*, by R. W. Pickford (London, 1951); *Occup. Psychol.*, **29**, 182–92; *Brit. J. physiol. Optics*, **14**, 2–26; *Annu. Rev. Psychol.* **8**, 1–28. Additional references will be given where necessary.

The lantern test has been widely employed, often in conjunction with other tests. It was over-complex, but a much better and more simple lantern has been devised by Giles—the Giles-Archer lantern. Another type of test using coloured lights was devised by Hering, and was based on his theory. It employed red, green, yellow and blue glass filters, through which daylight reflected by mirrors could be viewed.

The most popular of all colour-blindness tests have been those based on the pseudo-isochromatic principle. Numbers, letters or geometrical shapes are printed in such colours that the distinctions between figure and ground tend to disappear for the colour-blind. The best known are Stilling's Tables, the Ishihara Test, the American Optical Company's Plates, the Dvorine Test, the HRR Polychromatic Plates, and several others, including certain plates published by Kalmus and by Farnsworth for blue-vision defects. These tests have all been valuable, and they have improved the efficiency of testing and increased the proportion of major defectives readily detectable from about 5 per cent to about 7 or 8 per cent of men in our population. The position of pseudo-isochromatic tests remains essentially unaltered, however, after half a century of research. They are capable of acting as cut-out tests for major defectives, but for fine discrimination they are useless, and their validity is very low because they tell us almost nothing about kinds and degrees of defects.

In 1901 Rivers reported on Lovibond's tintometer as a test for measuring small differences in colour sensitivity, using it in a series of very interesting researches. Pierce showed in 1934 the need for tests using finely graded differences of hue and saturation, but his difficulty lay in standardising the graded disks which he employed. This difficulty has now been overcome in the Farnsworth-Munsell 100-Hue Test, which is a valuable instrument for research. The Farnsworth Dichotomous Test, which is a quick and reliable test using coloured caps of the Munsell hue series, serves to cut out major red-green defectives, and, since it involves comparisons within a graded series of various hues, it may also indicate yellow-blue weaknesses successfully. Among more elaborate tests, the Colour Aptitude Test, by F. L. Dimmick, is coming into use for industrial purposes.

The Rayleigh Equation and Anomaloscope Tests

The opportunity for practical colour-vision and colour-blindness testing to become an efficient branch of applied science was given by the invention of the anomaloscope. It was shown by Lord Rayleigh in 1881 that the colour match between a variable mixture of lithium red and thallium green on the one hand, and sodium yellow on the other, could be used as a test for degrees and types of red-green colour-vision defect. The Nagel anomaloscope is a spectrometer in which there are three slits, providing beams of red (670·8 mμ), green (546·0 mμ) and yellow (589·3 mμ) light. The first and second of these are combined and their slits are controlled in such a way that the light from one slit is reduced when that from the other is increased by varying the widths of the slits. The third light is a standard yellow of variable intensity for comparison. In one model of this instrument the telescope swings under the control of a micrometer screw, and many colour equations may be selected, but the particular relationship of lights chosen is always dependent on the fixed positions of the three slits. For those equations in the middle of the spectrum a desaturating white light must be superimposed on the standard colour. This can be done by reflection from the near face of the prism, with a Vierling white light attachment. For testing blue defects the equation of a mixture of bluish green (518·5 mμ) and indigo (464·5 mμ), with a standard of 'cyanide' blue (486·1 mμ) has been recommended. This seems to call for the use of the desaturating light if a match for saturation as well as for hue is to be obtained. The chromaticity diagram is somewhat curved between these points.

It is possible to carry out the Rayleigh equation with coloured papers. Rayleigh himself used colour filters in his original apparatus, and several types of anomaloscope have been devised, most of which are filter instruments. Five anomaloscopes in addition to the Nagel instrument were described by Willis and Farnsworth in 1952, and valuable comparative data were given on their use. A filter anomaloscope was made for Miss Agnes Crawford in collaboration with Messrs. Ilford, Ltd., and was described by her in 1951. A filter anomaloscope was made in this Laboratory and data about its use and standardisation, with instructions on the technique of testing, have been published. A very much improved model will soon be available. It uses filters of narrow wave-band, and the efficiency of a filter anomaloscope depends on the narrowness of the transmission bands of its filters and the wideness of their separation on the chromaticity chart. For the Rayleigh equation the Ilford 'Spectrum' red, green and yellow filters are excellent, and a yellow-blue test may be made with the Ilford 'Spectrum' yellow and blue filters and a neutral standard. The lights obtained with these filters are dim, of course, but this is not necessarily a disadvantage. Interference filters have narrower transmission bands and more intense lights, but Chance's optical-glass filters can be used readily, although they have wider bands. Tests of blue and green matched against a blue-green standard, and of red and blue against a standard of violet, are quite feasible, and the writer reported on their use in 1951. A test of yellow and blue against a neutral standard is much more satisfactory for practical purposes, and easier to carry out, although the red-blue and blue-green tests seem more reasonable on a trichromatic theory.

Results and Value of Anomaloscope Tests

Three important forms of variation of colour sensitivity emerge clearly as a result of anomaloscope testing. They are variations of (1) matching range, (2) mid-matching point, and (3) shortening of the red (or violet) end of the spectrum. The matching range falls between the limiting points at which the subject tested sees the variable mixture as being just not different from the standard in each direction, namely, just not redder or greener than the yellow, for the Rayleigh equation. Similar ranges may be obtained for other pairs of colours. The mid-point is the centre of the matching range, and therefore may be shifted with the matching range to one side or the other, or may be normally centred. This normal centre position is simply the average mid-matching point for the random population tested. In Rayleigh's and other early tests attention was drawn to the mid-points, because the subject was required to make a number of settings, say three, of which the mean was taken. It is more important to measure the matching range in the first place, because this determines the true position of the mid-point.

We may find a large or small matching range, coupled with a large or small deviation of mid-point, and these differences may be compared with results for a representative sample of the population being tested. This comparison may be used to measure the individual's peculiarities of colour vision on a standard test. Shortening of the red end of the spectrum is very important in testing dichromatic protanopes and protanomalous subjects. If a yellow-blue test is used, shortening of the blue-violet end of the spectrum will also be found in some individuals.

Increases of deviation of mid-point lead to what has been called 'deviant', and, when more extreme, to 'simple anomalous' colour vision. Increases of matching range lead to 'colour weakness', and, when more extreme, towards 'colour blindness' in the stricter sense, and ultimately to dichromatic colour vision. Combinations of large deviation and large matching range are characteristic of the subjects called 'extreme anomalous' by Franceschetti in 1928, while dichromatic colour

vision is found where the matching range extends to the limits of the scale and deviation is no longer a meaningful measurement. 'Deviant' and 'colour weak' vision may be called 'minor' variations. 'Simple anomalous', 'extreme anomalous' and 'dichromatic' colour vision may be called 'major' variations. In general all dichromats are failed by Holmgren's Wool Test, while all the groups of major defectives together are failed by the Ishihara and other pseudo-isochromatic tests.

The distribution of red-green mid-points is markedly tri-modal, a large peak corresponding to the normal and deviant population, and two small peaks, one on each side and separated from it, corresponding to the simple protanomalous and deuteranomalous subjects. The distribution of red-green matching ranges is strongly J-shaped, the bulk of the population being normal or nearly normal, and the extreme anomalous and dichromatic subjects being collected towards the other end of the scale, with few intermediates. Those who do fall between these limits are mostly the not very extreme deuteranomalous subjects.

These facts have been established by extensive anomaloscope testing. The cruder cut-out tests for what is popularly called 'colour blindness' all depend on dividing the population into two groups, the normal and minor defectives on the one hand, and the major defectives on the other. If large natural gaps in the distributions of defects did not exist, such tests would be useless.

If other pairs of lights, in addition to red and green, are used in the anomaloscope, such as orange and blue-green, yellow and blue, and yellow-green and violet, the results show that the variability tends to be collected into the red-green and yellow blue directions. No subjects are found who are abnormal in orange and blue-green or yellow-green and violet, who are not more abnormal in red and green and/or in yellow and blue. The same types of variation are found in yellow and blue as in red and green, but the extremer forms are very much more rare, while the intermediate forms are much more frequent. The extremer forms are called tritanopia (and tritanomaly) corresponding to protanopia (and protanomaly), and tetartanopia (and tetartanomaly), corresponding to deuteranopia (and deuteranomaly). It is important to say that a deviation or weakness in red-green vision does not necessarily imply the presence of a deviation or weakness in yellow-blue vision, or vice versa, but the two pairs of variable sensitivities are independent.

Correlations Between Matching Ranges and Deviations

In order to test the theory often accepted, that the greater the deviation the greater the matching range, the following product-moment correlations were calculated between matching range and deviation in red-green vision. For 31 men who had large green deviations and/or enlarged red-green matching ranges, $r = -\cdot065$, and for 21 similar women $r = +\cdot047$. For 57 men who had red deviations and/or enlarged red-green matching ranges, $r = -\cdot101$, and for 50 similar women $r = +\cdot201$. None of these correlations is significantly different from zero. For 17 protanomalous subjects, $r = -\cdot347$; for 26 extreme protanomalous, $r = \cdot1174$; for 56 deuteranomalous, $r = \cdot381$; and for 67 extreme deuteranomalous, $r = -\cdot459$. The correlations for both deutan groups are significant on the ·01 level. Those for protans are not significantly different from zero.

If all subjects with the modal red-green matching range or more, and with red or green deviations of one or more times sigma, are excluded, there still remain a considerable number of subjects who have deviations and/or enlarged matching ranges in yellow-blue vision. Correlations may be calculated for these subjects between matching range and deviation in yellow and blue vision, apart from any possible influence of variations in red and green sensitivity. Among subjects who had normal red-green vision, but blue-yellow deviations and/or enlarged matching

ranges in these colours, the following product-moment correlations were obtained: between yellow deviations and yellow-blue matching ranges, for 47 men, $r = \cdot 0697$, and for 43 women, $r = -\cdot 573$; between blue deviations and blue-yellow matching ranges, for 34 men, $r = -\cdot 406$, and for 55 women, $r = \cdot 0032$. The correlations between blue deviations and blue-yellow matching ranges for women, and between yellow deviations and yellow-blue matching ranges for men are not significantly different from zero. The others are significant but negative. Matching range does not go with deviation, either for red and green or for yellow and blue vision, and the theory is not supported.

From the results described it is apparent that the anomaloscope has a special value, namely that it may be used to detect and measure the variations from one individual to another in matching ranges and deviations, and this can be done conveniently for yellow and blue vision as well as for red and green. These are the essential directions in which colour vision varies, apart from shortening of the red (or violet) end of the spectrum, which can also be tested. Four-colour testing does not necessarily commit us to a colour vision of the Hering type, as will be explained later.

The Inheritance and Comparative Frequencies of Red-Green Blindness

Numerous pedigrees for colour blindness have been collected, and they have given very strong support to the theory that red-green blindness is a Mendelian sex-linked recessive. The theory is therefore generally accepted as true. The relative infrequency of women defectives is an interesting point which supports the theory. The expected frequency of women defectives for a sex-linked recessive of which the frequency among men is about 7·8 per cent, as in colour blindness, would be about 0·61 per cent. The observed frequency of women defectives, however, is usually lower than expected, a point which will be dealt with later. For example, in 1927 Waaler reported tests on 9,072 girls and 9,049 boys in Norway. He found 0·44 per cent of major defective girls and 8·01 per cent of major defective boys. In 1956 Kherumian, Baudry, Lacorne and Moullec reported testing 3,873 men and 3,735 women in Paris, and found 9·33 per cent of major defective men and 0·51 per cent of major defective women. Pickford reported in 1947 that he found 7·8 per cent of major defectives among 464 men and 0·65 per cent (namely three) major defectives among 460 women in Glasgow, but Brown reported finding 7·43 per cent among 525 boys and 0·46 per cent (namely one) among 252 girls in the West of Scotland in 1950. The differences between these percentages might be due to chances of sampling.

There are interesting differences in the frequencies of red-green blindness in various geographical areas and racial groups. Rivers tested 320 men and 183 women Todas, in 1901 and 1902, and found 41 men (12·8 per cent) and 2 women (1·1 per cent) to be red-green blind, using Holmgren's wools, Nagel's card test and Lovibond's tintometer. Vernon and Straker reported in 1943 that there were 7·49 per cent of defectives among 123,414 men tested for the Royal Navy with the Ishihara test and some plates of the Stilling Tables, and significantly more in the South and West of Britain, where hair, skin and eye pigmentation are greater, than in the North and East, where they are less. Professor P. E. Vernon tells me, however, that subsequent tests in which an anomaloscope was used on about 6,000 men, did not give much support to the earlier findings of differences according to district and pigmentation.[1] Pickford was not able to find any connection between frequencies of defect and pigmentation in the West of Scotland. Professor R. Kherumian, of Paris, tells me that he has been unable to find any significant

[1] Thanks are due to Dr. N. A. B. Wilson for permission to quote these data.

differences throughout the Departments of France, using several pseudo-isochromatic tests on 6,215 subjects.

Among white men the proportion of red-green major defectives is usually about 7 or 8 per cent, but in 200 Jewish men Garth found only 4·0 per cent.[1] The proportion was 5·07 per cent among 473 Turkish men. Clements found only 1·9 per cent of defectives among 620 American Indian men, and 3·71 per cent among 323 American negroes.[2] The lowest figure reported was 0·5 per cent (2 individuals) among 387 Fijian men by Geddes in 1946.

Colour vision is found in lower animals, such as insects, fish, reptiles and birds, but not among frogs and toads or the lower mammals. It appears to have been evolved again by the higher apes, and colour blindness may be found among them too. The lower frequencies of these defects among men of other than white or Caucasian stocks suggests that natural selection may have acted more forcefully on this character among them than among Whites. We have no knowledge, however, whether the defects may have varied in frequency in any populations within historical times, and linguistic data, which have often been quoted as evidence, are too doubtful to be of much help to us. If natural selection does act upon colour blindness, tending to eliminate it, this would probably be through recognition of enemies, and choice and selection of food substances. The ripeness or over-ripeness of fruits, and the degree of decay in other foods, are often indicated by colour changes not apparent to the red-green blind. In many animals sexual selection appears to depend partly on colour, and this is true of some of the higher apes, but apparently not true of man.[3]

Although it is generally agreed that red-green blindness is less frequent among Pakistanis, Indians, Chinese, Japanese, negroes, American Indians, and others of darker skin, than among Caucasian Whites, yellow-blue defects of a minor kind are more frequent among them. In a short research in Glasgow with the Ishihara test and anomaloscope, on 29 Indian and Pakistani men, 25 white men and 22 white women, Agha M. Akram found that the normal white men had significantly greater red-green matching ranges, while the normal Indo-Pakistani men had significantly greater yellow-blue ranges, thus supporting findings reported by Pickford in 1951. The numbers tested were very small, but he also found only two major red-green defectives among the Indo-Pakistani men, while three were found among the white men and none among the white women.

The Two-Locus Theory and Multiple Alleles

In his survey of Norwegian boys and girls, Waaler showed in 1927 that there were significantly fewer defective girls than would be expected on the theory of a single locus for a recessive sex-linked gene for colour blindness. The expected frequency of defective girls for an observed frequency of 8·01 per cent of defective boys would be more than 0·64 per cent. He found 0·44 per cent. This suggested that there might be two kinds of gene for colour blindness, namely one kind for protanopia and another for deuteranopia, and that when these two were present together in separate loci in a double heterozygote, each would be neutralised by its own dominant normal gene, and the colour vision would be apparently normal. The difference between the 0·64 per cent expected and 0·44 per cent observed would, on this theory, be made up of these double heterozygotes.

[1] Useful data were given by Dr. M. Collins in her Presidential Address to this Section: *Brit. Assoc. Ann. Reports* (1937), pp. 207–26. A report on results obtained all over the world is being prepared by Kherumian and Pickford.

[2] Clements, F. (1930): 'Racial Differences in Color-Blindness', *Amer. J. Phys. Anthrop.* **14**, 417–32.

[3] A valuable discussion of the rôle of colour vision in the evolution of the primates and of Man is to be found in Polyak, S., (1957): *The Vertebrate Visual System*, Chicago U.P., pp. 969–74.

Many single pedigrees have been found which support this theory. Kherumian and his colleagues estimated in 1956 that there should be 0·52 per cent of red-green defective women in Paris, on the two-locus theory, and this was in close agreement with their actual finding of 0·51 per cent. If the relative frequencies of defective women in various racial groups could be established satisfactorily, the theory could be tested more efficiently. For instance, among Jewish women we should expect significantly fewer than about 0·16 per cent, the square of the percentage for men. Among 175 Garth found none. For Turkish women fewer than about 0·26 per cent would be expected. Among 217 none were found. Negro women should include fewer than about 0·14 per cent, but among 722 Crooks found 0·14 per cent, which would accord with the single-locus theory. Some researches support the two-locus theory, when there are significantly fewer colour-blind women than would be expected on the single-locus theory, and others apparently do not. Unless we know more about the reliability and validity of the tests used, and about the actual proportions of the genes supposedly occupying the two different loci in the various populations tested, we cannot assess the value of the researches adequately.

Owing to the work of Lord Rayleigh and others it was realised that there are several forms of red-green major defects in addition to protanopia and deuteranopia. This knowledge led Franceschetti to put forward in 1928 the theory of multiple alleles for colour blindness. He suggested on the basis of anomaloscope tests that there might be two series of allelomorphic genes, the protanopic and the deuteranopic series, each with its special locus, its normal gene and three alternative genes producing different degrees or types of defect. He supposed that the greater the defect the more recessive its gene in each series.

Many pedigrees have supported this theory, such as those in which the father and daughter are deuteranomalous while the son is a dichromatic deuteranope and the mother is apparently normal, or variants of this pattern. Similar patterns have been found for the protanopic types of defect. Some pedigrees have supported the view that there are three intermediates in each series, namely, in the one series simple protanomaly without shortened red, simple protanomaly with shortened red, extreme protanomaly and dichromatic protanopia, and, in the other series simple deuteranomaly, extreme deuteranomaly with strong green deviation, extreme deuteranomaly without green deviation and dichromatic deuteranopia.

More support for the theory has been derived from a statistical study of the frequencies of the types of women defectives as compared with the corresponding frequencies among men. According to the theory, since two defective genes of the same series are required in a woman to produce a defective phenotype, and the lesser defective genes are dominant to the greater, it would be expected that there should be significantly fewer women in proportion than men who are dichromats or extreme anomalous than simple anomalous, just as there are fewer defective women than men altogether. This prediction was verified by observation in this Laboratory in 1956, but much more work needs to be carried out on the problem.

Among 204 major red-green defective men tested in Glasgow the following proportions of the different types were found:— 15 protanomalous, 31 extreme protanomalous and 21 protanopes; 54 deuteranomalous, 71 extreme deuteranomalous and 12 deuteranopes. Groups comprising some of these and some others, all of whom carried out the same anomaloscope test, are shown in Diagram I. This diagram shows red-green matching ranges and mid-points for 166 male and 25 female major red-green defectives on the Pickford anomaloscope. All were outright failures on the Ishihara test, and all had more than three times sigma of deviation, either to red or to green, and/or more than four times the modal matching range. Although there are about three times as many women defectives in proportion as would be expected by chance for the number of men defectives shown,

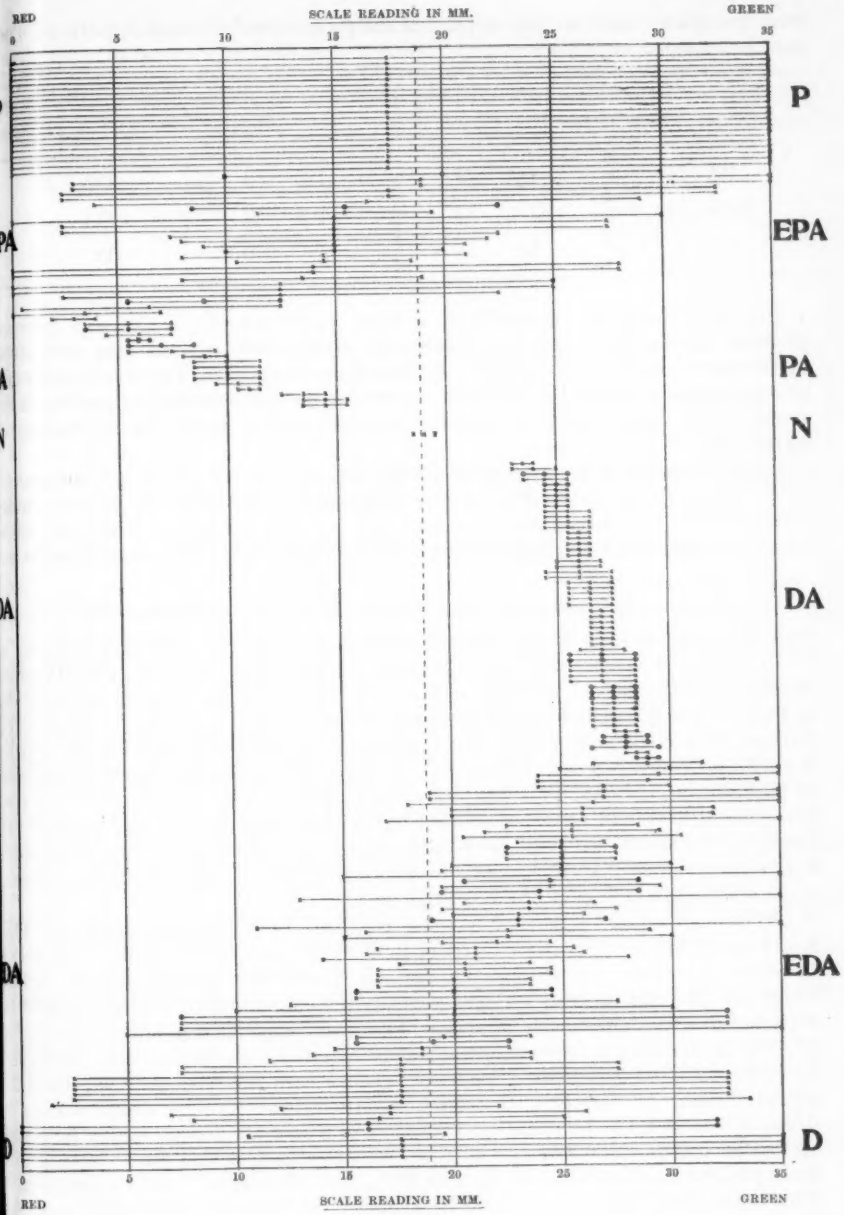


Diagram I.

Matching Ranges and Mid-Points for 191 Major Red-Green Defectives on the Pickford Anomaloscope.
One normal subject is shown in the middle for comparison.
P = protanope; EPA = extreme protanomalous; P = protanomalous; D = deuteranope; EDA = extreme deuteranomalous; D = deuteranomalous; N = normal.
Crosses = men; circles = women.

both groups are fair random samples of the populations of major defectives. They may be grouped as follows:—

Men (crosses)				Women (circles)			
PA	14	DA	45	PA	3	DA	11
EPA	24	EDA	58	EPA	2	EDA	9
P	21	D	4	P	0	D	0

For these numbers of men and women major defectives, the gap between protans and deutans should be filled with about 2,000 normal men and about 5,500 normal women. The graph therefore shows merely the top and bottom tails of a very extended distribution. One normal subject is shown for comparison in the middle. The mean yellow match for normal subjects is at a scale reading of 18·85 mm.

It is to be noted that a number of EDA subjects have red deviations and several EPA subjects have green deviations. Willis and Farnsworth obtained a similar result (1952). The distinction between EDA and DA subjects is where deviation and matching range both begin to decrease, and similarly for EPA and PA subjects.[1]

Colour Blindness in Turner's and Klinefelter's Syndromes

Anomaloscope testing, which reveals types and degrees of colour blindness, is being used by the writer in conjunction with Dr. John S. S. Stewart, of the Western Infirmary, Glasgow. In conditions in which there is a sex reversal because the individual has developed organs contrary to the 'genetic sex', the frequency and pattern of inheritance of colour blindness should be that expected for the genetic sex, that is, for the XX or XY chromosome condition, rather than for the masculine or feminine phenotypes. This was suggested by Polani. The 'genetic sex', it is now generally agreed, can be demonstrated by the technique of 'nuclear sexing'. All tissues in the female have at least 50 per cent of cells which show ' sex chromatin' believed to represent the XX constitution. In the male less than 4 per cent of cells show anything which could be mistaken for this.

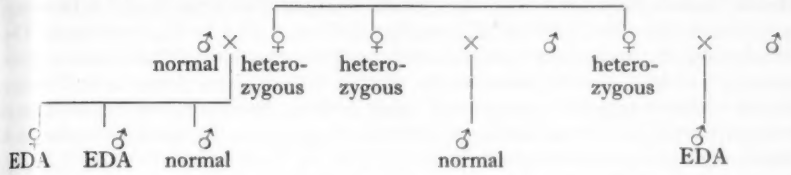
Dr. Stewart tested 24 'genetic females' who had the Klinefelter syndrome. This is a form of sub-fertility which occurs in subjects who look like men but about two-fifths of whom are genetic females. He found that all had normal colour vision on the Ishihara test, whereas about two colour-blind individuals would be expected in a random sample of about 24 ordinary men. In a group of 26 subjects affected by the Klinefelter syndrome, all of whom were genetic males, however, he found four red-green blind, and this accords with the frequency expected for ordinary men. In Turner's syndrome, a form of sub-fertility which occurs in subjects who look like females, but almost all of whom are genetic males, he found one to be colour blind among the first two subjects tested. Polani, Lessop and Bishop found 4 colour-blind subjects in 25 females affected by Turner's syndrome.[2] This corresponds with what would be expected for ordinary men rather than for ordinary women.

In this work the anomaloscope is especially useful because it enables us to show that the type of colour defect of a given subject is the same as or different from that

[1] Thanks are due to Mr. R. Lakowski, Mr. A. M. Akram and Mr. R. Brown for help in collecting some of the data, and to Mr. T. McMenamay for making the diagram.

[2] Polani, P. E., Lessop, M. H., and Bishop, P. M. F. (1956): 'Colour Blindness in "Ovarian Dysplasia" (Gonadal Dysplasia)', *Lancet*, **2**, 118–20.

of any colour-defective relatives, and this makes the interpretation of the pedigrees much more decisive. The following unusual pedigree for extreme deuteranomaly (EDA) in a woman shows how the inheritance followed the pattern for a man, presumably because she was a genetic male of the type found in Turner's syndrome. The extreme deuteranomaly was closely similar in the affected brother



and sister, and their mother showed a minor defect according strongly with the view that she was a heterozygote, whereas the father's colour vision was normal. The mother had two sisters, both of whom showed minor defects suggesting that they were heterozygotes, and the son of one of them was normal in colour vision. The son of the other sister showed extreme deuteranomaly, again closely similar in type to that of the woman affected by Turner's syndrome, although he was a spastic with a speech defect which made testing a little difficult.

Other Points about the Inheritance of Colour-Vision Defects

There are several other interesting points about the inheritance of colour-vision defects, and the first is that statistical evidence shows that major defects are, in general, incompletely recessive. This was first suggested for protanopia by Schmidt, who showed that nine protanopic heterozygotes all had a characteristic loss of luminosity at the red end of the spectrum (Schmidt's sign). It was shown by Pickford in 1951 that women known to have relatives who had red-green major defects were significantly more often slightly defective in red-green vision themselves than other women. In addition, if these women were classified according to the kinds of defects found in their relatives, their own colour vision varied in minor ways in accordance with their relative's major defects. In a recent test of the parents of a totally colour-blind girl of the so-called rod-vision type, the mother had general colour weaknesses of a marked degree and the father also had weaknesses but less marked. The parents of a totally colour-blind man of the same type had similar defects. The minor defects were much more frequent in the four parents than in the population at large. This suggests the possibility that achromatopsia of this type is incompletely recessive.[1]

A second point concerns minor variations of colour vision in general. These consist of small deviations to the red, green, yellow and blue sides of the colour equations used as tests, and small increases of matching range. The inheritance of these variations is not known, but each type is continuously distributed, although the distribution of minor matching ranges is strongly skewed, and they might be due to small groups of minor genes which have additive and subtractive effects. Minor red-green variations are discontinuous in distribution from major red-green defects, but minor yellow-blue variations appear to be continuous in distribution with extreme variations of yellow-blue vision. Accordingly, tritanomaly and tritanopia, tetartanomaly and tetartanopia, might be the limits of very extended continuous distributions, and the great rarity of these defects in comparison with the relative frequency of major red-green defects accords with this hypothesis. Nevertheless, tritanopia might be due to a dominant autosomal gene, as suggested by Kalmus in 1955, and tritanomaly might be a sex-linked recessive. Total colour blindness of the so-called rod-vision type might be an autosomal or a partially

[1] *Nature* (1957), **180**, 926–7.

sex-linked recessive, the defective gene being located on the homologous portions of the X and Y chromosomes.

Another interesting point, probably concerned with pigmentation, is that three unrelated albinos, one woman and two men, have been tested in Glasgow, and all were found to have defective red-green vision. That of the men approximated to extreme protanomaly, that is to say, much enlarged matching range with a very big red deviation, but without shortening of the red end of the spectrum. This kind of colour defect has not been reported in any other subject. That of the woman, however, was like simple protanomaly, that is to say, very large red deviation without enlarged matching range, and again without shortening of the red end of the spectrum. The woman failed on the Ishihara test, but the men both passed it. All had slight yellow-blue defects.[1]

Theories of Colour Blindness

In order to account for variations in colour vision reported in previous pages, it seems that not less than three general hypotheses would be needed. These three have all been put forward in one form or another before, but the present review will serve to focus attention upon them afresh.

The three hypotheses are:—(1) the "loss" hypothesis, (2) the "collapse" hypothesis and (3) the "swing" hypothesis. The "loss" hypothesis assumes that some variations or defects of colour vision might be due to the absence of cones with certain functions, to the absence of their retinal connecting fibres, or to lack of one or more of the normal light absorbing pigments of the retina. The "collapse" hypothesis assumes that other defects or variations might be due to the presence in the same cones of light-sensitive pigments normally found in different cones, or to the abnormal linkage of cones which act separately in normal vision. The 'swing' hypothesis assumes that still other defects or variations might be due to some abnormality of the light responding system as a result of which its spectral sensitivity is changed. It is assumed that defects and variations occurring in any of these ways may be combined and may occur in complete as well as in partial forms.

Rushton's recent work on the differential absorption of light reflected from the interior of the living human eye has brilliantly vindicated the loss and collapse hypotheses, so far as dichromatic colour vision is concerned.[2] His protanopes lacked the red-absorbing pigment, but both the red- and the green-absorbing pigments were found in his deuteranopes' eyes. The theories that protanopes have a loss defect and deuteranopes a collapse defect seem therefore to be substantiated at least for the small number of subjects actually tested, but it remains an interesting question whether the collapse defects of deuteranopes are due to abnormal linkages of cones which should respond separately, or to the presence of both pigments in all red and green responding cells. It is also a question whether some dichromats do not have extreme swing defects, as will be mentioned later, and whether some deuteranopes do not have loss defects.[3]

When we turn to anomalous subjects the position becomes more complex. If all degrees of protanomalous defect were due to varying proportions of loss, then we should expect that the greater the loss the greater the matching range and shortening of the red end of the spectrum, but no red deviation. Progressive loss would be a progressive 'filter' effect, and would influence the yellow produced by mixing red and green in the same way as a spectral yellow. Shortening of the red end of the spectrum and red deviation are not correlated, but matching range is not correlated with them either. No doubt the experiments upon which this

[1] *Nature* (1958), **181**, 361–2.

[2] Rushton, W. A. H., Human Cone Pigments, Paper 1, *Symposium on Visual Problems of Colour*, Nat. Phys. Lab., Teddington, September 1957.

[3] Graham, C. H. and Hsia, Y. (1958), Color Defect and Color Theory, *Science*, **127**, 675.

assertion is based should be repeated with better equipment, and it is intended that this should be done, but there is no expectation that there will be any change in the conclusion. In those subjects who show the lack of correlation most clearly, namely the simple protanomalous subjects in whom the matching range is almost normal and the red end of the spectrum is not shortened, although there is a red deviation in the Rayleigh equation of six or seven times sigma, it seems clear that we cannot be dealing with a loss or collapse defect. Many normal and extreme anomalous subjects present us with the same difficulty in less sharply defined ways.

In the various degrees of deuteranomaly, if we had to deal only with appropriate degrees of red-green collapse, it would be expected that there should result corresponding increases of matching range without any deviation. Certainly it is true that all degrees of matching range without deviation can be found, but in most cases there is a considerable green deviation, and occasionally a slight red deviation, just as there is sometimes a green deviation in protanomalous subjects, and the typical cases of simple deuteranomaly have deviations as large as six or eight times sigma, coupled with little or no increase of matching range. It seems that we cannot be dealing with collapse defects at all in some kinds of deuteranomalous colour vision, and that in other kinds they must be present in combination with other defects.

It is clear, therefore, that in many, but not in all examples of protanomalous and deuteranomalous colour vision, we must be dealing with kinds of defects other than could be accounted for on the loss or collapse hypotheses alone. The swing hypothesis applies to these cases, and, if it is vindicated, its effects must be capable of being combined with loss and collapse defects in a variety of ways, and of appearing in any degree of magnitude.

Since there is seldom any correlation between matching range and deviation in red-green vision, whatever is said of major defects applies to minor red-green variations with appropriate changes. Since there is either no correlation or a small and negative correlation between matching range and deviation in yellow-blue vision, after all subjects with red-green variations have been excluded, what is said of red-green defects and variations applies with appropriate changes to yellow-blue variations and defects as well.

According to Schmidt, whose work is based on that of Trendelenburg, further distinctions must be made for extreme anomalous subjects. Those who have enlarged matching ranges with the eye adapted to neutral light and also with undefined adaptation are called colour 'amblyopes', while those who have enlarged matching ranges only with undefined light adaptation are called colour 'asthenopes'. The more recent Nagel anomaloscopes are provided with a standard adapting light to test these differences, which present further problems for a complete theory of colour blindness. It would appear that the colour amblyope has both collapse and swing defects even after neutral or undefined adaptation, while the colour asthenope has both kinds of defects after undefined adaptation but loses the collapse defect as a result of adaptation to neutral light.

There has always been a possibility that some deuteranopic defects might be due to loss of green-responding cones, substances or retinal connections, and it was claimed by Abney in 1895 that in deuteranopes the green part of the spectrum would have lowered luminosity. Graham, Hsia and Berger reported in 1955 a reduction in sensitivity to wavelengths below 515 mμ in some deuteranopes, including the deuteranopic eye of one woman with unilateral deuteranopia. This might support a trichromatic theory, but could fit equally well into a four-colour theory. Most workers have not reported such a loss, and it may not apply to all deuteranopes. A comparable loss has not been reported for yellow sensitivity in tetartanopes, minor yellow-blue defectives with blue of normal luminosity. Deuteranopic and tetartanopic defects have therefore been regarded as wholly collapse defects.

We must ask what is the probable nature of swing defects. We may say that loss defects are due to the absence of certain photochemical substances, or perhaps to the failure owing to some other cause in the transmission of certain signals. Collapse defects arise because certain cones respond equally to two ranges of light-waves which are normally differentiated, so that, for example, red and green rays give rise to the same sensation. Swing defects, however, must imply that certain cones, which normally respond to particular rays and give rise to what we accept as the appropriate sensations because we regard ourselves as normal and have these as our usual experiences, continue to respond to these rays but do not give rise to the normal sensations. Those with marked swing defects must, in Houstoun's words, be "colour different" rather than colour blind. In loss defects in red, for example, the red signals must be missing at least in part, and there is no red sensation if the loss is complete. In red-green collapse defects the red and green signals must be confused, either partly or completely, because appropriate mechanisms respond equally to either kind of light-ray. In red swing defects, however, there must be a shift in the red-sensitive pigment in the short wave direction, without necessarily shortening the red end of the spectrum, and in green swing defects there must be a shift in the green-sensitive pigment in the long-wave direction. In these ways it would be possible to have the conditions observed in deviants and anomalous subjects who have deviations, and, presumably, all deviations of mid-matching point in the Rayleigh Equation or its equivalent for other colours must be caused by swing effects. Although deviation and matching range are not positively correlated, however, some swing defects might be so extreme as to be additional causes of dichromacy.

Three- and Four-Colour Theories

While the trichromatic theory holds the attention of the majority of workers, there are aspects of colour vision and colour blindness which have always suggested a four-colour opponents theory. Spectral response peaks for such a theory suggested by Jameson and Hurvich in 1955 and 1956, on the basis of their observations, would be: red (with two peaks) at 620 mμ and 440 mμ, yellow at 586 mμ, green at 490 mμ and blue at 467 mμ, for one observer. For another observer red was the same, but the other three were: yellow at 580 mμ, green at 500 mμ and blue at 475 mμ.

In their work on the quantitative aspects of a provisional four-colour opponents theory in 1955 and 1956,[1] Jameson and Hurvich have assumed that there would be only three photochemical pigments, but that they are involved in two pairs of response processes for hue, and that the same three pigments would be involved in the response system signalling luminosity. 'Red' would be signalled by the first and third pigments, and 'green' as an opponent to it by the second alone. 'Yellow' would be signalled by the first and second, and 'blue' by the third alone as an opponent to 'yellow'. 'White' would involve all three, and 'black' only the recovery reactions. It is possible, therefore, that experiments to test the number of photochemical pigments in human colour vision might appear to support the trichromatic theory, whereas in fact the mode of operation of the pigments might be in accordance with a four-colour theory with white as a separate variable and two pairs of opponents for hue discrimination. Jameson and Hurvich have made a valuable analysis of the phenomena of colour vision, colour blindness and anomalous colour vision, showing that many of the difficulties might be met by a four-colour opponents theory of this type. Their detailed quantitative treatment, with predictions which can be checked against experimental results, is particularly valuable.[2]

[1] Hurvich, L. M., and Jameson, D. (1956). *Trans. N.Y. Acad. Sci.*, **2**, 33–8.

[2] Jameson, D., and Hurvich, L. M. (1956), *J. Opt. Soc. Amer.*, **46**, 1075: and (1957) Further Developments of Quantified Opponent-Colour Theory, Paper 22, *Symposium on Visual Problems of Colour*, Nat. Phys. Lab., Teddington.

A trichromatic theory, however, based on three cone pigments with appropriate modifications of spectral response functions where necessary, would be sufficient to account for all the forms of abnormality and variation described here. The application of the trichromatic theory to protanopia, protanomaly and red deviations, to deuteranopia, deuteranomaly and green deviations, and to enlarged matching ranges in 'normal' subjects, has been explained. In yellow-blue vision it might be assumed that blue deviations are due to small shifts in the blue-sensitive pigment in the long-wave direction, while yellow deviations are due to shifts in the same pigment in the short wave direction. Enlarged matching ranges in 'normal' yellow-blue vision, and in extremer forms of yellow-blue defect, might be due either to losses of blue-sensitive pigment or receptors, giving a shortening of the blue-violet end of the spectrum, or to collapse of discriminating power between the blue-responding cones on the one hand, and the red and green cones taken together on the other hand. Increases of blue-yellow matching range with losses of luminosity in the blue-violet end of the spectrum might also be due to pigmentation of the transparent parts of the optical system, which tends to increase with age. The idea that all deviations in normal and anomalous colour vision, and even some forms of dichromacy, might be due to so-called 'shifts' in the spectral sensitivity of certain pigments, suggests that there must be several different pigments grouped under each of three headings, red, green and blue. The detailed investigation of such possibilities by physiologists would be of the greatest interest to psychologists. It seems worth suggesting that the possibility of a number of photosensitive pigments grouped under three main headings might link up with the interesting suggestion of Stiles that there might be as many as five photosensitive mechanisms in cone vision.[1]

Colour Vision Defects and Industry

An important linkage between science and industry is to be found in colour-vision research. It is well known that from some occupations, in which life and safety depend on the use of amber, green and red signal lights, all major colour-vision defectives must be excluded, and that the less seriously defective among them may be the more dangerous because they do not appreciate their own weaknesses. In other occupations fine colour matching is required, as in the colour printing, the dyeing and the paint industries, or discrimination is required between near shades of desaturated grey, blue, yellow, green, pink and so on, as in the work of some telephone electricians. Not only must the major defectives be excluded from such industries, but there will be a level of sensitivity required among normal subjects, characteristic of each industrial task, to be found only by experimental research. Testing yellow-blue vision may be important, especially since it tends to deteriorate somewhat in persons over about forty-five years of age. R. Lakowski is engaged in a research on this problem now. For all such purposes a simple anomaloscope would have much higher reliability and validity than any other tests in common use. It is essential, however, to use very narrow wave-band filters, even if they give only a dim light, or spectral lights if they can be obtained.

In conclusion it is interesting that research has recently been directed by Sloan towards trying to find signal lights which would be at least fairly satisfactory for colour-vision-defective subjects. Other research by Bally has dealt with the relationship between the magnitude of colour-vision defects and the degree of liability to accidents in ordinary road traffic. Such practical applications of psychological testing for colour blindness should be developed much further.

[1] Stiles, W. S. (1953): Further Studies of Visual Mechanisms by the Two-colour Threshold Method; *Union Internationale de Physique pure et appliquée, Coloquio sobre Problemas Opticos de la Vision*, Madrid.

THE CAPE AS AN ANCIENT AFRICAN FLORA[1]

ADDRESS BY PROFESSOR R. S. ADAMSON
PRESIDENT OF SECTION K

HAVING spent the greater part of my active life as a botanist in Africa it is natural that I should turn to that continent for the subject of this address.

Africa has often been referred to as the 'Dark Continent', but from a botanical point of view this will soon be a misnomer. The flora of the continent is undergoing rather intensive study and very soon few parts will be without the basis of a standard flora. For example, floras have been published or are in course of publication for Algiers, Morocco, Egypt, tropical West Africa, East Africa and Angola, and are in active preparation for the Rhodesian Federation and Mozambique. Plans are well under way for South Africa also.

All this activity is at present taxonomic. In no part of Africa has there been attained the degree of intensive study that is associated with modern floras of the north temperate zone. Cytological or genetical investigations are few and scattered and palynology is still in its infancy. Even distribution data are in many cases very far from complete. As a result, any phytogeographical conclusions must be rather tentative and less secure than elsewhere.

Africa is a large continent with much variety in conditions and with varied flora. The continent is divided into two very unequal portions by the huge desert area of the Sahara. To the north of this is a small strip to the north-west on the Mediterranean coast which has a flora that is commonly regarded as being more allied to that of Europe than to the rest of the continent. Good, for example, places this flora in the Boreal Kingdom. The remainder of the continent is placed in the Palaeotropical Kingdom, except for the flora of the south-western Cape, which forms a separate kingdom. It is with this last flora that I propose to deal.

Though this Cape flora occupies a very small part of the continent, it is in its typical form exceedingly distinctive and very sharply separated from the others. So much so is this the case that botanists familiar with the floras of central Africa find this one quite strange.

This flora contains a high percentage of plants not found in other parts of the continent. It is also exceedingly rich in species, as rich indeed as any part outside the tropics. For example, the very small area of the Cape Peninsula, 180 square miles, possesses over 2,600 species. Contrast this with South Australia with 2,468 species for an area of 380,000 square miles, or with 2,400 species for the whole of Great Britain. This richness is expressed in several ways. The communities of the vegetation rarely exhibit dominance of one or a few species. Instead they are built up by a number with close similarity in life-form and leaf character. Dominance of species seems confined to the earlier phases of development, either at the drier margins of the region or in communities developing after destruction by fire or other interference. The floristic richness is also shown in the large number of species with limited geographical range. Allied species replace one another in adjacent areas.

With so rich a flora it is difficult to give shortly the diagnostic features. Perhaps the most striking feature is the abundance of Proteaceae, Restionaceae, Ericaceae, and woody Composites, along with abundance of Iridaceae and Cyperaceae. The last are most often not marsh or water-demanding plants. Grasses individually are

[1] Presidential Address delivered to Section K (Botany) on Thursday, August 28, 1958, at the Glasgow Meeting of the British Association.

abundant but never approach dominance. Grasslands are quite wanting. As compared with northern floras there is a complete absence of catkin trees with the single exception of Salix, and a paucity of such families as Ranunculaceae, Cruciferae, Caryophyllaceae or Rosaceae. On the other hand, families scarce in the north are here abundant, such as Aizoaceae, Santalaceae or Thymelaeaceae. The flora has been the source of a number of well-known garden plants; foremost are the very numerous species of Pelargonium. Others are Nemesia, Gladiolus and other Iridaceae, many annual composites, and recently many species of the aggregate Mesembryanthemum. The white arum lily is also worthy of mention.

An analysis of this rich flora can be made by grouping the species in accordance with their geographical ranges. On so doing, there appears a high percentage of plants not found elsewhere, endemics. Among such endemics are whole families such as Bruniaceae, Penaeaceae, Grubbiaceae and others, also many genera and very numerous species. A rough calculation from Phillip's *Genera of South African Plants* shows that about 35 per cent of the genera are endemic. Of these only about a quarter are monotypic. In all the larger families there are endemic genera, e.g. in Compositae 60, Iridaceae 20, Aizoaceae 17, Ericaceae 15, Papilionatae 14, Proteaceae 11, and Restionaceae 9.

Endemic species are exceedingly numerous and occur in almost all genera. Some are widespread in the area of the flora, some are exceedingly restricted, and there are all intermediate stages. As an example, in Erica there are 102 species recorded from the small area of the Cape peninsula. Practically all are endemic to the area of the Cape flora but nineteen are confined to the peninsula itself.

The richness of the flora and the large number of endemics show that the region has been one favourable to species differentiation. Among the endemics are some that are to be regarded as relic and ancient species. Such often exhibit some discontinuity in their range. Some of these have been regarded as among the more primitive in their genera. Others do not exhibit such features, and though often much limited geographically, appear to be relatively recent differentiations.

A second component in this flora consists of plants which have their main area of distribution in the southern hemisphere. Among these are many of the most characteristic parts of the flora and include such families as Aizoaceae, Cunoniaceae, Haemodoraceae, Proteaceae and Restionaceae. In many of these families the African representatives are distinct from those in other southern lands. There are also a large number of southern genera belonging to widespread families. Examples are Caesia, Cliffortia, Danthonia and its allies, Malvastrum, Pelargonium, Phylica and Tetragonia. While in Africa all these genera have their main area in the Cape region, several do occur outside.

These two components, endemic and southern, make up a large and most characteristic portion of the flora, but there is a large component of plants found both in this flora and in other parts of Africa but not in the other southern lands. This is not a homogeneous group and is divisible into subgroups. In a first, there are those plants with their main area in the Cape region, but which also occur outside. Examples here are Aristea, Muraltia, Stoebe and many others. Erica belongs here. Most of the plants in this group have a range outside the Cape area that is discontinuous and in some cases there are very large gaps. These are certainly to be looked on as characteristic members of the Cape flora. Associated with these is a group of herbaceous plants, in Africa mostly confined to higher altitudes, which also occur in the north temperate zone. Several are markedly discontinuous in distribution. Examples are seen in Anemone, Alchemilla, Hieracium, Viola or Valeriana.

Another subgroup is composed of plants belonging to this flora but which are rather widespread in other parts of Africa. The plants here are less strictly confined to mountains than those in the previous subgroups. Examples out of many are:

Gladiolus, Hermannia, Euclea, Selago and Rhus. A majority of the trees composing the forests associated with this flora are in this category.

A fourth subgroup is made up of plants which have their main area of distribution in central Africa and which occur here rather as outliers. Examples are Asclepiadaceae, Aloe, Polygala, and the tribes Andropogoneae and Paniceae among the grasses. Both these subgroups might be considered as intrusive elements in the flora.

A final component is composed of plants with a world-wide range. These, which are not numerous, are either species of cosmopolitan genera or water or marsh plants of what are termed followers of man. Nowadays there are of course numerous alien plants which need not be considered at the moment. Apart from annuals, most of the aliens that have become completely established are of Australian origin.

The territory occupied by this Cape flora consists of a series of mountain ranges that run more or less parallel with the coast and are separated from it by a rather narrow coastal plain. These mountains reach a height of 3,000–7,000 ft. and in their upper parts have a climate that has a higher rainfall, greater humidity, and less prolonged dry periods than the lower parts. In spite of this climatic contrast there is a marked uniformity of life-form from sea-level to the highest summits. There is an altitudinal zonation in species but not in life-forms.

There is reason to suppose that the area at present occupied by this flora is smaller than it was in the past. On its inland margin it is bounded by the arid region of the Karroo which has a quite distinct flora. The southern part of the Karroo is a broken plateau with numerous small hills and some mountain ranges. Those mountains near the main area carry the Cape flora in its characteristic form though with a number of local endemics. On the higher of the detached hills there are often found patches, large or small, of characteristic Cape species. These occur mostly on sheltered south-facing upper slopes. They can be traced as far as the escarpment of the upper Karroo. Such patches, which form islands surrounded by the arid karroo flora, are confined to those sites with the least arid conditions. At one time they must have been continuous with the main area. Now they are relics left in a region that has become increasingly arid.

While this effect of increasing aridity is clearly shown at the margins, it is also traceable in the main area of the flora. In many genera of the flora of the mountains it has been found that there is a maximum concentration of species at or about that part where the east-west and north-south ranges meet, with a decrease in numbers both north and east. This is such a general feature that Levyns has put forward the suggestion that it is a criterion for the true members of the flora and that those plants which do not exhibit it are really intrusive members of another flora. I do not propose to enter on any discussion of this but merely to point out that this region of maximum species concentration is that with the highest rainfall and least probability of prolonged drought. If the climate as a whole is becoming drier this would be the part least affected, and the part where survival of less tolerant species would be most likely. In support it has been found that in a number of genera, these species regarded as least specialised are confined to mountain-tops. Evolution in the flora seems to have been in the direction of increasing xerophily.

The mountains are an old and stable habitat, but the coastal plain, which is either a dissected peneplain or sandy flats, is geologically relatively recent. It has a drier climate than the mountains, from which its flora in the main must have been derived. This plains flora contains a number of endemic species. Many of the plants characteristic of it do not show the distribution pattern noted above. The flora as a whole appears as a derivative from that of the mountains which has become adjusted to drier conditions. In it are most of those plants noted as outliers from central Africa.

The flora of the karroo on the inland boundary of the Cape region is quite distinct and one evidently derived from that of the other parts of South Africa.

Along the southern margin, however, there is some evidence of derivation from Cape flora. Examples can be seen in some of the xerophytic species of Pelargonium, Tetragonia and others. Some of the geophytes that occur in the southern karroo area may be survivors of the Cape-like flora that once covered the area.

The above signs of retreat in face of a drying climate mark this flora as an old one, a characteristic further supported by its clear distinction from adjacent floras, by its great richness in species, its high percentage of endemics, and its general lack of species dominance. But though an old flora it is evidently one in which differentiation of species has been active, as is shown by the large numbers of local endemics and the number of genera with numerous species.

Though at the present time this flora has its main development in the small area in the south-western Cape, there are to be found patches or traces of it in many parts of the continent. The flora is one so distinctive floristically that its elements can be traced even when occurring in vegetation of a quite different type.

On the mountains of the eastern Cape there is a zone of quite characteristic Cape flora which occurs above the altitudinal limit of forest. Further north, in the Drakensberg and on the higher mountains as far as Abyssinia, there are patches of vegetation commonly referred to as 'heath' which are essentially similar. This gets its name from an abundance of Ericaceae. It is developed in places with a cool moist climate at the higher altitudes. Its composition and structure vary considerably from place to place but there is in all cases a notable proportion of typical Cape plants. These are associated with and sometimes dominated by plants more characteristic of warmer conditions at lower altitudes. For example, on the eastern mountain border of Southern Rhodesia, the vegetation was dominated by tussock grasses between which occurred numbers of Cape plants. Among the characteristic Cape plants in this 'heath' are Erica, Philippia, Blaeria, Stoebe, Helichrysum, and more locally Anthospermum, Cliffortia, Restio and others, together with herbaceous plants such as Aristea, Romulea, Sebaea, Wahlenbergia and Xyris. There is also often a considerable proportion of those herbaceous plants noted earlier as occurring on the mountains and in the north temperate zone.

The 'heath' vegetation, which is a mixture of elements of the Cape flora and of the flora of warmer regions, does not form at all a continuous zone but as separated patches with intervals which are sometimes large. While some of the intervals are the result of lack of suitable habitats, not all can be so explained. Discontinuities in the distribution of the individual plants in the vegetation are even more marked. As a whole this 'heath' seems best looked upon as being made up of relics of the Cape flora which have become mixed with plants from the lower levels. The existing relics would seem to be parts of what must have once been a continuous area.

Along the west coast the existence of outliers of the Cape flora or its constituents is less apparent. North of the Kamiesberg in Namaqualand is a wide stretch of desert country in which Cape plants can find no possible habitat. However, on the mountains further north there are some limited patches in which Cape plants characteristic of the drier parts of that flora are present. There are also some isolated occurrences of Cape plants in Angola and even as far as West Africa.

The large continental island of Madagascar lying off the east coast is separated from the mainland by the Mozambique channel which is 200–300 miles across. The island has a flora with a quite unusually high proportion of endemics. Without entering on any details, the island flora can be separated into two parts. That on the east coast is tropical and more closely related to those of the lands surrounding the Indian Ocean than to that of most of Africa. On the mountains and on the drier west side there is a different flora which especially on the higher land has a definite affinity with that of the Cape, or rather that the 'heath' of the mountains. There are many genera in common and many close parallels. Among the genera are

Aristea, Caesia, Chironia, Hypertelis, Lebeckia, Philippia, Restio and Stoebe. Proteaceae are represented on the island by an endemic genus. The similarity in flora is such that Professor Humbert, on a visit to the summit of Table Mountain, remarked that he felt he might be on a peak in Madagascar. Though almost all the species are different, there is a very clear affinity with the Cape flora.

Turning now to the other end of the continent, the vegetation of the strip of country on the Mediterranean coast in the north-west has for long been compared ecologically with that of the Cape. There is general similarity in structure and in the distribution of life-forms. In composition, however, the flora at first sight appears to be so different that as noted it is placed in a separate floral kingdom. Analysis brings out that this is the result of a very large European component composed of plants not found elsewhere in Africa. Associated with this are a number of endemics. There is, however, a distinct African component, though less numerous than the European. In the African component are a number of those plants which occur in the Cape flora though not confined to it. Well-known examples are: Ornithogalum, Romulea, Gladiolus, Lotononis, Psoralea, Lycium, Wahlenbergia and others. Erica, though occurring on the north side of the Mediterranean, is lacking on the African coast.

In addition to these African plants, there is an interesting group of plants found in this flora and in that of the Cape but absent for the most part in any intervening region. Though many of these are neither conspicuous nor abundant in the flora their distribution is important. Examples of such plants are: Androcymbium, Aizoon, Chenolea, Cytinus, Erodium, Laurentia and Leyssera. Along with these are to be placed a number of plants which though distinct have their nearest allies in the Cape flora. Some of these seem quite unrelated to any northern plants. In nearly all there is discontinuity with their southern relations. Perhaps the best known example is Calendula, which is allied to the South African Dimorphotheca and belongs to a tribe otherwise wholly southern. Other examples are Echium and Lobostemon, Sedum and Crassula, Tetraclinis and Widdringtonia, Globulariaceae and Selaginaceae, and possibly Iris and Moraea.

The above indications will be sufficient to show that this Mediterranean flora, which at first sight seems so different, has affinities with that of the Cape. The plants giving this affinity very often show discontinuities in distribution and would seem to be an old part of the flora. It is a less numerous part than that of the European component, which would appear as a more recent one which is partially replacing the older one.

That there is such a relationship is supported by a study of the floras of the Atlantic islands, Madeira and the Canaries. These islands are volcanic and do not seem ever to have been part of the continent. Madeira is about 350 miles from the mainland, the Canaries 60–70 miles. The island floras have always been regarded as allied to and in all probability derived from that of the Mediterranean region. They differ especially in a rather high proportion of endemics. In the floras there is a similar association of components as was noted for the mainland. That part which is allied to the Cape flora is here larger. It is shown in a number of genera in common and in several which are allied though distinct. Some of the endemic genera are more nearly related to Cape genera than to any others. Examples are seen in Anagyris, which is nearest to the Cape Podalyria, and Bencomia to Cliffortia. In some large widespread genera, species endemic to the islands are more closely related to Cape species than to those of North Africa.

The two small volcanic islands, St. Helena and Ascension, can receive very small notice. Both lie in the South Atlantic, St. Helena 1,200 miles from the mainland, Ascension 800 miles. On both islands the original indigenous flora has been very much reduced and largely replaced by aliens. Only scraps now remain to indicate what it once was. So far as it is known, however, this indigenous flora was very

largely endemic but clearly allied to the Cape flora. Of twenty-eight indigenous genera on St. Helena, no fewer than twenty-two are represented in the Cape flora. Of the endemic genera, too, some are nearly related to Cape genera: for example, Nesiota, which is closely allied to Phylica which itself occurs on the island.

There is little doubt that the floras are closely related. The poverty of the island flora can be attributed to the distance from the mainland and the great difficulties of transport.

From the foregoing very brief sketch it emerges that all these floras have affinities with that of the Cape and that all seem old, at any rate in those components that produce these affinities. Their present-day widely separated distribution on the continent and islands is most readily explained on the assumption that there was once continuity between them and that they are now relics of a once widespread flora. That there was such connection between the parts is supported by the existence of Cape plants in various parts of Africa in vegetation and associated with a flora quite unlike that of the Cape. Protea is a well-known example. While the old flora was evidently widespread, it is not suggested that at any time it covered the whole of Africa. It must at all times have been restricted to those parts where the climate was temperate and without prolonged periods of drought. Such conditions might be expected at the extremities of the continent, on the islands, and on high ground which may well have been much more extensive than it is today.

The present-day wide separation of the fragments points to the time when they were continuous as being a remote one. There must have been a very long period during which retreat or disappearance has been taking place. At the period when there was continuity it is reasonable to suppose that the flora was less varied than it is anywhere today. There has been continued species evolution in the Cape region and to some extent in all the areas. The existing Cape flora is not to be looked on as typical of the original flora but as a modified descendant from it. Such continued evolution, along with the lack of uniformity of dispersal to be expected in an area as large as that of Africa, goes a long way to explain the dissimilarities in the existing fragments. Thus, as a general proposition, it is suggested that in the past there was a widespread temperate flora that covered all parts with a suitable climate, which later became very much reduced in extent and the parts somewhat changed by species evolution.

Such a proposition goes a long way to resolve a number of much discussed phytogeographical difficulties. For example, arguments as to whether the separated patches or traces represent relics of a migration from south to north or vice versa fall away. One or two examples will clarify this. The genus Erica has a few species on the mountains in central Africa, some in western and south-western Europe, and about 400 in the Cape region. This does not imply a centre of origin in the Cape nor a survival there of an early southward migration, though both have been put forward. It seems more probable that this was an early genus which had a wide dispersal in the old flora and one which has found conditions especially favourable to differentiation in the Cape region. Exactly the same argument can be applied to Stoebe, in which there are a few species on the African mountains and a larger number at the Cape, the latter being looked upon as the more highly evolved and presumably more recent. A recent discussion on the distribution of the African Proteaceae points in the same direction. In this family, two or possibly three genera have a wide and decidedly discontinuous distribution through the continent. The remaining eleven genera are confined to the Cape region. The former would appear as direct descendants of the old flora, while some of the Cape genera may be later differentiations.

The present distribution and floristic features of the fragments indicate an early date for the break-up of continuity and a long period of isolated development. As noted, most of the portions have their own endemics. Some possible indication of

the length of this time may be obtained from the fact that the highly endemic flora of Madagascar has such clear affinities with that of the Cape. These are such that a land connection must have existed when the flora was first established. The geological structure of the island shows that such connection with the mainland did exist in the past. Geologists, however, are not in complete agreement as to the date when severance occurred. It has been variously assigned to a period as early as the Upper Cretaceous to one as late as the Pliocene. Whatever the final decision, there is no doubt that the flora must have been established and widespread before the separation; also, that since separation there has been time to allow of the development of the very highly endemic flora of the island.

The break-up of the once continuous area of the old flora and its disappearance in many places, is attributable to climatic changes and especially to changes in the direction of aridity. A reduction in the extent of highland with suitable climate would also be a contributory factor. It is known from a variety of sources that there have been climatic fluctuations in Africa. Wet and dry periods, cooler and warmer, have alternated. Wetter and cooler times would have been favourable to the spread of the flora, drier and hotter ones would have been adverse. The latter seem to have been the more pronounced. For example, in the Miocene and Pliocene, desert and steppe conditions were widespread. At such times this flora must have become confined to those localities on highlands or near the coast where conditions were less severe. In the Cape region itself, the mountains which provide the main habitat have been in existence without serious change since the beginning of the Tertiary epoch. Even in the driest periods it is probable that the higher summits retained a climate sufficiently moist to permit continuance of the flora. While major changes towards aridity were destructive to the flora, minor ones would be favourable to species differentiation.

It is noteworthy that the floras associated with the warmer regions appear to have been more adaptable in the occupation of arid or semi-arid regions. As an example, in the Pliocene, rather extreme desert conditions existed in the Kalahari and extended over much of Southern Rhodesia and the northern Transvaal. This region has since had a great amelioration in climate but its present flora is one derived from the warmer parts with very few traces of the old Cape flora. At a later period arid conditions developed further south in the region now occupied by the Karroo. The great part of the Karroo flora is certainly derived from the north, though, as noted earlier, there are some traces of the old flora.

The one-time continuity of the fragments of the old Cape flora must have been before the onset of the arid conditions of the middle Tertiary, that is, it must have been in the early Tertiary or in Cretaceous times. At this period Africa was separated from the northern lands by an ocean, the Tethys Sea, which extended from Gibraltar, to the Indian Ocean. In Jurassic and early Cretaceous times this sea covered most of what is now the Sahara. There were, however, a number of islands in the north-west, and it would seem that the old flora must have become established on these. Such conditions would explain the relative poverty of Cape plants here. Later, in Miocene or Pliocene times, when aridity of climate was causing great reduction in the old flora, a land connection was established across the straits of Gibraltar. This permitted invasion from the north into a region already isolated and reduced by a drying of the climate. It seems that in this way the existing flora with its great preponderance of the European component came into being.

So far consideration of the distribution of this old flora has been confined to Africa, but it has been recognised for a long time that there are floras which exhibit striking similarities to that of the Cape in south-west Australia and in Chile. In both regions the floras contain a large number of the southern plants so characteristic of the Cape. Perhaps the closest similarities exist between the floras of the Cape and of south-west Australia. Both have the same structure of the vegetation. Floristically

both are characterised by the presence of Proteaceae, Restionaceae and other southern families. There are also a large number of genera in common, and many cases where a southern genus in one flora is represented by a closely allied one in the other. Examples are seen in Phylica in Africa and Cryptandra or Pomaderris in Australia, in Tetraria and Gahnia and many other pairs which could be quoted. The Australian flora has many endemics but they do not form such a high percentage as in the Cape flora.

There are, of course, a number of striking differences in the floras. The proportions of several families are quite different. In Proteaceae there are no genera in common. In Restionaceae, which are much less prominent in Australia, there are only two genera in common, both of which have been divided by some authors. Families characteristic of the Cape flora such as Ericaceae or Oxalidaceae are wanting in Australia, while those characteristic of Australia such as Epacridaceae or Dilleniaceae are absent in Africa. Again, in several families the representatives in the two floras belong to different tribes: Rutaceae and Orchidaceae are striking examples.

Though such differences might be much expanded, there seems little doubt that they are outweighed by the similarities which point to a rather close affinity, an affinity indeed as close or even closer than that between some of the portions in Africa.

If similar comparisons are made between the Cape flora and that of South America the same kind of relationships are revealed, though rather less close. It is to be noted that there is a greater similarity between the floras of South America and Australia than between either and the Cape. Also that the floras of the subantarctic islands are much closer to that of South America than to that of Africa.

These similarities, and especially the large southern component common to these floras, make it difficult to avoid the conclusion that even though separated so widely geographically they had a common origin and that they represent today fragments of a one-time single floral area. In any case their present distribution and possible origin call for some explanation.

In pre-angiosperm times, Permian, Triassic and early Jurassic, these southern lands had a very distinctive fossil flora which was rather uniform wherever it has been preserved. This fossil flora has been recorded from rocks in South Africa, Australia, South America, and also in India and in Antarctica. The fossil flora was not only uniform but also distinct in many features from contemporary floras in the lands north of the Tethys Sea. In the southern lands this fossil flora had such uniformity that land connections uniting the regions where it has been found have been postulated. The resulting more or less united land mass has been called Gondwanaland.

For present purposes the importance of this Gondwanaland is that it provides clear evidence that evolution took place south of the Tethys Sea along lines distinct from those occurring in northern lands. Additional support for this is obtainable from the characters and distribution of the Gymnosperms of the southern and northern hemispheres. The older angiosperms also must have had similar distinct lines of evolution in the south and in the north.

As we have seen that the original Cape flora existed before the time of the diminution of the Tethys Sea, it is reasonable to picture it as having arisen in the southern lands. Such an origin leads to possible explanations of the present distribution. In general, there have been two schools of thought on this. The first, for long regarded as the more orthodox, pictures the seat of origin of the angiosperms as having been in the tropics or even north of the equator. From thence a series of migrations, with accompanying modification, led to the floras of more temperate lands and of those with less favourable habitat conditions. On such a view the Cape flora would represent the end phase of an early southward migration to a temperate climate, the last traces of which persist at or near the extremities of the continents. Hypotheses

of this nature have the advantage that they provide an explanation of present distribution without the necessity for the postulation of any large changes in the distribution of land and sea. They ignore the importance to the old separation of northern and southern lands at the period when angiosperms began to appear. They also fail to give an explanation of other features such as the greater similarity between the floras of Australia and South America than of either with that of Africa. Many of the characteristic features of distribution of the southern genera are also left vague.

The second school of thought is one which has been given an increasing degree of support in recent years. This is that the temperate flora itself arose in the southern lands, either in the cooler south or on higher ground elsewhere. Views of this nature have even been extended to the suggestion that all angiosperms arose in the south. Such hypotheses are, however, beyond the present scope.

The most obvious difficulty facing all hypotheses of southern origin lies in the great discontinuity of land in the southern hemisphere and the apparent need to postulate enormous disappearances of former land connections. Such difficulties are met to a large extent under the theory of 'Continental Drift'. According to this, in early Mesozoic times Gondwanaland was a more or less continuous land mass. South America is thought to have approached or even been joined to Africa and to have been continuous with Australasia by way of Antarctica, which was then further from the pole. India is supposed to have extended much further south and to have been connected with the other lands.

This continuous mass of Gondwanaland is supposed to have begun to break up in Jurassic or early Cretaceous times. A first phase was the separation of southern Africa. South America is pictured as remaining for a time in connection with Australasia to the south. This connection became broken by a southward drift of Antarctica and an eastward one of Australia.

If one can accept this hypothesis and its time scale, there is no doubt that it provides the simplest and most reasonable explanation of the facts of distribution. It would, for example, explain at once the uniformity of the various parts of the southern fossil flora. Though it is outside the present purpose, it gives the most satisfactory explanation of the observed facts in relation to the southern ice-age which occurred in Carboniferous to Permian times.

On the immediate purpose, the early temperate flora that exists today in the Cape flora and its associates is to be pictured as having become established over the more temperate parts of a continuous Gondwanaland. When this became broken up, the flora persisted in those parts of the resultant lands where the climate was suitable. Subsequent to the separation of the lands, each of the areas of the flora would have been exposed to vicissitudes of climate and to possibilities of invasion from surrounding regions. All these factors, isolation, climatic change, and invasion, would tend towards local changes from the original flora.

In the Cape itself some traces of invasions can be recognised. Parts of the African component are plants characteristic of warmer and moister regions which presumably entered during one of the wetter phases. Many of the forest trees would be included here. On the other hand, there are a number of karroo-like plants, especially succulents, such as Aloe, Cotyledon and members of the old genus Mesembryanthemum, which seem to have entered during one of the drier phases. As near the coast such plants never attain dominance, and as the drier mountain slopes, the coastal plain and much of the marginal zone of the flora, have today a flora clearly derived from that of the mountains, it would appear that here the climatic changes were never severe. There is a striking contrast between these coastal parts and the conditions on the inland margin where extensive disappearance of the Cape flora is apparent.

The problem of the origin of this old temperate flora is very much a matter of speculation. It seems most reasonable to picture its having attained its general

features in lands south of the Tethys Sea. But there is no reason to look on it as being among the very early angiosperm floras. Its diversity, and the presence in it of so many widespread families and genera, rather point to its having originated from some already diversified flora.

The view has been put forward recently that in spite of the absence of fossil evidence, there were already in existence angiosperms at the time that the Gondwanaland fossil flora was flourishing. Further, that these were already differentiated into a warmer forest flora and a more temperate one. The flora we have been considering on such a view would have originated in the more temperate old flora, though in the course of time it would have received additions from adjacent warmer regions. The large southern component in the flora might be regarded as the direct descendant of the original flora.

To sum up, the distinctive Cape flora appears as part of an ancient temperate flora that at one time occupied the cooler portions of the southern lands. This original flora would seem to have itself arisen in the southern hemisphere. At present it exists in a number of separated fragments each of which has undergone modification as the result of local differentiation and of invasion.

The Cape region itself, the largest of the remaining fragment in Africa, has been a region of active species differentiation. Though the flora is ancient, it is one that shows evidence of modification in relation to altered conditions, drying of the climate, and it cannot by any means be regarded as moribund.

ADLARD & SON LTD.
BARTHOLOMEW PRESS, DORKING, SURREY
Telephone : Dorking 3874
PRINTERS TO SCIENTIFIC SOCIETIES
Experienced in mathematical formulae composition

THE PARTNERSHIP BETWEEN INDUSTRY AND EDUCATION[1]

ADDRESS BY SIR WILLIS JACKSON, F.R.S.

PRESIDENT OF SECTION L

IT is the practice of the Committee of the Section to arrange for the Presidency to rotate among representatives of the several branches of educational activity, and this year the honour has fallen to a representative of industry. This recognition of industry's place in our educational structure is of great importance. I greatly welcome the opportunity it affords me not only to pay tribute to the very considerable service to education which is being performed by many industrial concerns, but also to make forthright comment on the fact that many others are making a much smaller contribution that is demanded both by their own best interests and by the well-being of the country as a whole.

To some extent differences of attitude within industry on educational matters are understandable. Industry is a highly diverse organism—in the size of its component parts and in the nature of its activities. As shown in Appendix I, there are over 190,000 manufacturing concerns employing less than 100 people compared with 79 employing more than 5,000. Among these are firms the activities of which are constantly being refreshed and revitalised by application of the results of scientific discovery and technical development, and which recognise, without persuasion from outside, that the further education and effective training of their young recruits are vital to their continued progress. There are others the basic science of whose activities is unclarified and in which there is little urge to search for, and then to apply it; and where what was found, or considered, to be educationally adequate for their past recruits is still not thought to require modification and supplementation.

We have good reason to be grateful for the achievements of the former firms for upon them rests largely our ability to sustain the export business on which the provision of an adequate supply of foodstuffs and raw materials depends. And we have much cause to be anxious about the latter ones for their deficiencies may well undermine the prosperity which we hope to continue to enjoy.

A great deal has been said and written during recent years about our national shortage of scientists and technologists, and the urgent need for a larger number of our most able boys and girls to devote themselves to a prolonged study of science and its applications. The scale of the increased annual output required within the next decade has been assessed; steps are now being taken within the Universities and larger Technical Colleges to cater for it, and, as shown by the Advanced Level G.C.E. data of Appendix II, the Grammar and Public Schools are responding magnificently. Though one may regret that this situation was not achieved earlier in the event it gives much cause for satisfaction. It shows in particular that industry has gone some way towards establishing within the schools a belief that it can offer to those with a sense of vocation as satisfying a career as the 'learned' professions which have traditionally attracted a dominant proportion of our most able young people. I persuade myself that this means much more than a mere recognition of the material importance of scientific and technological achievement, and more than a growing acceptance that scientific subjects, properly taught, are no more specialised than, and provide as good a vehicle of education as, arts subjects. It means that industry is coming to be seen not merely in terms of machines and of detailed

[1] Address delivered on August 29, 1958, at the Glasgow Meeting of the British Association.

scientific and technical processes, but also in terms of men of character and stature, and that it provides great opportunities not only for the exercise of scientific and technical skill but also for co-operation in service and in human relations.

Yet I incline to the view that in stressing our shortage of scientists and technologists we have made inadequate recognition of the importance of the supporting contribution, and given too little attention to the educational needs, of the much larger body of our technicians and skilled craftsmen. All too infrequently is reference made to the fact that without their skill and resource no new idea can be translated from the laboratory and design office into an efficient and reliable piece of engineering equipment. The satisfactory education and training of technicians and craftsmen is no less vital to the successful operation of industry and to the health of the national economy than that of scientists and technologists, and they seem to me to raise the more difficult problems. These problems require the more urgent attention because of the increasing number of school leavers—see Appendix III—which will become available for employment during the next few years.

It would not be appropriate for me to discuss here the many issues between management and the trade unions which are inevitably associated with the introduction of technical change in industry. But one thing is clear—technical change normally alters the balance of the labour force; it raises the numerical importance of the scientific, technological and technical staff, and of the skilled men engaged on service and maintenance, relative to direct labour; and it requires of everyone, increasing versatility, adaptability and capacity for understanding changing situations and points of view. It demands, in fact, a higher level of scientific and technical knowledge and attainment in an increasing proportion of industrial personnel, and extended and broader general education for all. For the attainment of these objectives industry carries a considerable educational responsibility.

The task is of course a co-operative one between the Schools, the Technical Colleges, the Universities and Industry. Unfortunately this needs only to be stated for one to be conscious of the inadequacies of the present contacts and collaboration between—and indeed within—these constituent parts of our educational system. I cannot hope to touch on more than a part of this very large problem, and I therefore propose to concentrate mainly on what industry's contribution to its solution should be.

I have chosen as the title of my address 'The Partnership between Industry and Education'. The crucial word is partnership. I intend this as a stronger word than co-operation, for my concept of partnership is that each partner shall interest itself in, and, whenever appropriate, make a contribution to the work of the other partners, as well as fulfilling the responsibilities which accrue to it directly within its own domain. I will now try to analyse these two aspects of industry's responsibility.

The Schools

First with respect to the Schools. During the past few years many industrial concerns, stimulated by their desire to ensure to themselves an adequate recruitment of promising young people, have been seeking to establish closer contact with the Secondary Schools, and with this has been associated an attempt by many schoolmasters to acquire a fuller appreciation of the nature and achievements of industry, of its significance in the national economy and of the career prospects which it affords. This progress has been achieved in a variety of ways supplementary to the immense contribution of the Youth Employment Service—by vacation schools within industry for schoolmasters and senior students; conferences representative of industry and the schools organised under the auspices of the Federation of British Industries, the British Association for Commercial and Industrial Education, scientific societies and the professional engineering institutions; the preparation of films depicting industrial development—such as 'The Enquiring

Mind' by the Institution of Electrical Engineers—and their presentation in schools; the issue of special career publications by the national press and the distribution of descriptive training booklets by individual firms.

But the outstanding action of industry affecting the schools was the establishment in 1955 of the Industrial Fund for the Advancement of Scientific Education in Schools. The object of this Fund was to make capital grants for science buildings and apparatus within the Independent and Direct Grant Boys' and Girls' Schools which cannot receive assistance for capital works from public funds. At the inception of the Fund the then Minister of Education undertook to give special attention to the analogous needs of the maintained schools. Close upon £3·2 million was subscribed by 140 firms. Within less than two years this money had been allocated among 187 schools as building grants and 328 as apparatus grants, and many of the new facilities are already in use. The grant will produce in these schools a net increase of some 50 per cent in the accommodation, and much improved conditions, for science teaching, and make possible a substantial increase in the number of boys and girls studying physics and chemistry to the Advanced Level of the General Certificate of Education.

By these various means a much closer relationship and understanding has been established between those who are moulding the thinking of young people and those responsible for providing the opportunity for them to pursue satisfying careers. If the aim on the industrial side has been to ensure that more of the best students devoted themselves to the advanced study of science rather than arts subjects, there has been explicitly associated with it a desire to avoid a degree of specialisation on scientific study such as might frustrate the achievement of a broad general education. That the Executive Committee and the Assessors of the Industrial Fund had more in mind than a mere improvement of facilities and increase in numbers is evident from the following extract from their report.

> If, indeed, a big impetus has been given to science teaching, and the number of science specialists increases, the point may be reached when nearly two-thirds of the VIth will be studying Physics and Chemistry. There are some schools already with a larger proportion than this, but it would be surprising—and perhaps not desirable—if a larger proportion than two-thirds (other than exceptionally) was reached. The trend is still upwards and it presents Science Masters with a real challenge, for it means that already many of the ablest boys in the school who were formerly channelled into the humanities, are now turning to science. It is not enough to make researchers and technologists of them. At their best the humane studies in the VIth produced first-rate people of great flexibility of mind, who were in the best sense highly educated and cultured people. It is now the business of the science masters to see that their ministrations do not produce narrow specialists, but help the best minds with whom they are in contact to make of science the core of an equally liberal education. It can be done, and is done in some schools already, and it should be the goal of all schools in their dealings with science 'Specialists'. Any subject in the curriculum can be either arid or fruitful in its effect on adolescent minds, and in those schools which the Fund has substantially helped, the science staff now have their opportunity to realise the full potentialities of their subject.

I would add that this realisation must embrace the liberalising of the education of arts 'specialists', by the presentation, in a manner appropriate to them, of the methods, achievements and needs of science and technology, with reference particularly to the manner in which scientific and technological progress affects communal life and conditions increasingly the conduct of national and international affairs. I hasten to concede that this is much easier to advocate than to achieve; but also to emphasise that its achievement is of great importance.

Outside general comments of this kind industrialists are wisely hesitant about commenting on the development of secondary education. Yet I feel I must express an opinion on the technical subjects which form a part of many Secondary School, including some Secondary Grammar School, curricula. Broadly speaking these are

subjects, such as engineering science, geometrical drawing, elementary surveying, building construction, accounting, commerce, and economic geography, which have a recognisable association with industry or commerce. No doubt these subjects serve to direct some pupils' interests towards particular careers, and, because of their vocational significance, to provide an incentive to careful study. But valuable though this may be it does not by itself constitute an adequate educational justification. The justification for their inclusion must rest on the extent to which there is established in the student's mind an association between them and other subjects of the curriculum. The test must be the value of their contribution to general, not vocational, education.

There can be no justification for including technical drawing, for example, in the work of any secondary school if the aim is to be the production of a series of standard drawings by the use of projection constructions carried out by rule-of-thumb methods, with an idea of providing a pre-employment training and selection of draughtsmen. Yet many existing courses do little to suggest that they aim any higher than this. If this is the result of industrial request, then I would state emphatically that such requests are ill considered and misguided, and must be fought. Industry should not require of its juvenile recruits that they shall have received at school any specific vocational preparation for the occupations they are about to enter; and there can be no greater misunderstanding of the real needs of industry than to suppose that it should. This is not to argue against the Secondary Technical School. If, in the above case, the pupil is led to approach the subject from the need in craft work to 'describe' accurately a three-dimensional object by means of two-dimensional drawings, and to regard the techniques as a 'language' developed from mathematics for this purpose, there are excellent opportunities for co-ordinated study of craft work, science, mathematics and English. It is extremely important to aim at such co-ordination within the limits of ability of the students if real progress is to be made, and to avoid presentations which are narrowly factual, and which make little attempt to gauge and develop the capacity of students to understand as well as to learn.

It would be a serious matter if Grammar Schools were to regard technical subjects as representing a field of sixth-form study in which pupils of no great ability could succeed though they showed little aptitude for the traditional subjects of grammar school curricula, and if, with engineering as objective, boys were to be encouraged to remain to age eighteen for this purpose. Unless capable of undertaking advanced work in physics and mathematics with fair promise of success at Advanced G.C.E. level it is far preferable they should leave to enter industry at an earlier age and to pursue their studies within the Technical Colleges.

Further Education—The Technical Colleges

In respect of Further Education the responsibility of industry is much more direct since, in general, the students are now in employment. Prior to the late war the only means open to the vast majority of young workers who wished to acquire the knowledge and qualifications which they saw to be necessary for their progress in industry was attendance at evening classes. These classes remain a very important, and indeed a still growing, part of further education, but by themselves they are no longer adequate to provide for the technical needs of developing industry, let alone for the needs of the individual for continued general education. There has been a gratifying increase since the war in the willingness of industry to release its young employees to attend part-time day technical courses. The number of such students increased from 41,560 in 1938 to 414,160 in 1957, but very wide discrepancies exist between different branches of industry in the granting of this facility and there is much scope for further improvement. In its White Paper on Technical Education of February, 1956, the Government set 700,000 as its target

for the number of part-time day students, and it is to be hoped that this figure, at least, will be achieved.

Appendix IV shows the scale of the development of Ordinary and Higher National Certificate courses since 1938. These courses are now largely operated on a part-time day basis and this has reduced the wastage from the early years which was previously associated with evening study. But it is still very large. The basic reason is that many of the boys admitted to the early years are unsuited to this type of course. They enter with their eye on the fact that National Certificate courses have been a major route to exemption from the examination requirements of the professional Institutions, and are inclined to attach little value to other courses, much more suited to many of them, which lead to technician and craft qualifications of the City and Guilds of London Institute and similar examining bodies. So much so that, having failed, too many withdraw completely from further study. No doubt the situation would be improved by a clarification of the structure and of the inter-relationship of courses at the Preliminary stage. But basically the problem is psychological, and its solution demands of industry a much clearer recognition of the value which it attaches to success in technician and craft courses, and closer collaboration with the Colleges in the encouragement of its apprentices to participate in them.

The normal duration of day-time release from employment is one day per week, and it is necessary to consider whether this period is adequate. This is perhaps best discussed in the first instance in relation to students who aim to achieve professional engineering status via National Certificate courses. On the basis of one day plus some evening work in College per week, satisfaction of the examination requirements of the engineering Institutions now involves a minimum of eight years' study from entry in the first year of the Ordinary course. Even then it is open to question whether the sequence is well suited to, and the breadth of the course adequate for, professional purposes. Moreover, devotion of the whole period of College attendance to technical study affords no opportunity for participation in corporate life within the College and no time for the continuance of general education—now referred to as the liberalisation of technical education—except in so far as good teachers may achieve something of this within the technical subjects themselves.

These limitations may be overcome by the 'sandwich' arrangements which are being developed in connection with the new Diploma in Technology courses administered by the National (Hives) Council for Technological Awards. The normal structure of these courses is one of alternate periods of six months in industry and in College, with entry based on Advanced Level G.C.E. passes in Mathematics and Physics or Chemistry, plus four passes at the Ordinary G.C.E. level, or an average mark of 60–70 per cent in the Ordinary National Certificate examination. The larger firms of the engineering industry particularly are giving strong support to these courses and are financing the participation of their 'Student Apprentice' candidates throughout them. It seems likely that there will also be an increasing provision of similarly organised courses with a somewhat lower standard of entry leading to Higher National Diplomas. These developments suggest that Higher National Certificate courses may soon come to serve mainly the needs of high-grade technicians—as was the original intention—while still affording a path, albeit an arduous one, to professional status.

An important Hives Council requirement of courses submitted to it for recognition for the Diploma in Technology award is that they must have a substantial content of liberal studies. The initiative recently taken by the Institution of Electrical Engineers to stimulate the introduction of sandwich courses leading to a Higher National Technician's Diploma is a step in the same direction for the better education of this increasingly important category of industrial personnel.

And it must not stop there. It will be remembered that the Education Act of 1944 advocated the establishment of county colleges for the purpose of encouraging the continued general education of young workers in a form suitably related to their vocational interests. It must be regretted by many that this part of the Act has not been implemented, and that progress has been restricted to the organisation by a number of large firms of Works Schools, recognised and supported by their respective Local Authorities and the Ministry of Education, in which, within the limits of their capabilities and resources, they have attempted to fulfil the basic objectives of the plan.

These far-sighted attempts are highly commendable. They afford no basis, however, for a comprehensive solution of the problem of continued education beyond the statutory school leaving age for the large body of young people whose employment on repetitive operations will make little technical demand on them and who are likely to have increasing opportunities for the misuse of leisure. No aspect of further education calls for more careful thought and action than this one, and the sympathetic support of industry is vital.

In its solution it will require an expansion of educational facilities well beyond those already approved for the further development of technical education outlined in the White Paper of 1956, including of course an adequate supply of well-qualified and well-trained teachers. This supply is the most crucial aspect of all our educational plans.

The scale of the requirements of the Technical Colleges for full-time and part-time teachers during the next few years, the conditions of service which should be established for them, and the nature and organisation of their training, were reported upon by a Special Committee of the Ministry of Education in May of last year.[1] The recommendations of this Committee have been so widely approved that I will comment only on certain contributions from industry which are essential for their implementation.

In the first place more industrial concerns must take steps to ensure that full-time members of College staffs maintain close contact with their activities. This contact may take the form of research conducted in collaboration with them, of consulting work undertaken on their behalf, of periodic secondment to them for further experience, or of assistance with supervision of the practical training of sandwich course students during the works parts of the course. The initiative should no doubt desirably come from the teacher, but it will not be sustained in the absence of a willing, indeed an enthusiastic, response from industry.

In the second place, notwithstanding the contemplated reduction in the contribution of part-time relative to full-time teachers, the absolute amount of this contribution must increase. The composition of the part-time teaching staff in the Technical Colleges during the session 1955–1956 given in Appendix V shows that 60 per cent of it consisted of persons engaged in the professions, industry and commerce. These men and women help to ensure that the teaching work of the colleges takes proper account of day-to-day industrial and commercial practice and experience and reflects the trends of development. With the growth of day courses their assistance during the day time, and not only in the evenings, becomes increasingly necessary. Under these circumstances they will become much more closely identified with the full-time staff and with the corporate life of the Colleges than is possible with evening work. Indeed it is to be hoped that the Colleges of Advanced Technology and the Regional Colleges will find means of according to a small number of them special recognition and the opportunity of participating not only in the advanced teaching work but also in the formulation of academic policy.

The proposal of the Special Committee for the establishment of a Staff College in which senior staff members of the Technical Colleges would reside for periods

[1] *The Supply and Training of Teachers for Technical Colleges*. H.M. Stationery Office.

of a few weeks with representatives of industry and commerce for the purpose of identifying and studying together major problems concerned with the aims and growth of technical education, and with a view to stimulating new methods and attitudes within the colleges, would provide a further means of strengthening the link between the two sides.

There is still another way which is by no means yet fully developed or exploited. Strong Governing Bodies and Advisory Committees including active—not merely paper—representatives of the highest ranks of industry are essential if the concept of partnership is to achieve reality.

Industry Itself

I now come to a consideration of the responsibility which falls to industry within itself—that of ensuring that its juvenile recruits are given a sound practical training appropriate to the duties they are to perform, and not merely in the short term, following its completion.

An earlier President, and a long-standing supporter, of Section L, Sir Arthur Fleming, used often to say that the most valuable raw material of industry is its young people, and in his own organisation he planned accordingly. Nowadays many more employers give recognition to this important fact than was the case when he first began to expound it over forty years ago. Yet the dearth of satisfactory training facilities in industry is quite certainly the outstanding weakness of our present educational arrangements, and unless corrected as a matter of urgency will frustrate the developments now taking place in our Schools, Technical Colleges and Universities at a time when much increased numbers of young people are becoming available for employment.

It is well known that some industrial concerns devote much care and attention, and a considerable proportion of their resources, to the organisation of comprehensive training schemes. These schemes embrace *Graduate Apprenticeship* of up to two years' duration for holders of university degrees, with the variant that students who have been accepted by a university may spend a year in industry between leaving school and commencing their degree courses; *Student Apprenticeship* of four to five years' duration, which, as already mentioned, is coming to take the form of alternate periods of six months or thereabouts spent in industry and in a Technical College on a sandwich course leading to the Diploma in Technology or to the Higher National Diploma; *Technician Apprenticeship* of four to five years' duration, the entry standard for which may extend from an Advanced Level G.C.E. in one science subject, combined with a good Ordinary Level record, to a modest performance at the latter level, and in which the educational objective may be the Higher National Certificate or a Technician Certificate of the City and Guilds of London Institute; and *Craft Apprenticeship* to which probationary entry may occur between minimum school leaving age and 16 years, and which then extends from age 16 to 21. An important feature of such comprehensive schemes is the opportunity provided for promising boys to move upwards from one grade of apprenticeship to another.

These schemes are mainly associated with large firms which normally find it beneficial to provide apprentice training schools in which the early stages of training are organised and supervised by instructors wholly devoted to this purpose. They are not the only firms, however, which need men who have received a broad and thorough training. Indeed if the small and medium-sized firms of which British industry is largely composed do not have such men, and at all levels, they are unlikely to be able to withstand the pressures of some other industrial countries, particularly in the event of the establishment of a European Free Trade Area. So far they have managed by buying, when particular needs arose, the trained products of other firms. While in the interests of individual advancement no one should

question the desirability of reasonable mobility from firm to firm, the enormous scale on which it has occurred during recent years has been highly inequitable and, moreover, an extremely wasteful process. A more important consideration for the future, however, is the fact that the responsibility for training the increasingly large national pool of skilled men and women required to accelerate the technical changes, and to achieve the enhanced productivity, which are now possible cannot be carried by the few firms which had had the foresight to provide good training facilities for themselves. Additional comprehensive training facilities must be made available, and these can be achieved if the smaller and more specialised firms will collaborate in the organisation of group schemes in which their limited individual resources are co-ordinated and amalgamated. One such scheme is the recently introduced Scottish Electrical Training Scheme covering five electrical firms and two Electricity Boards, which is dealing initially with only professional training, and another, the Engineering Industries Association Group Apprenticeship Scheme concerned mainly so far with craft, technician and student apprentices.

I do not underestimate the difficulties which will be associated with a large-scale expansion of good training facilities, and it seems to me quite evident that they will develop too slowly if left to the initiative of industry alone. In any case the time has come when the national interest requires independent assessment of the adequacy of industrial training arrangements. It is remarkable that notwithstanding the careful supervision and control which is exercised throughout our educational system over the organisation of teaching, the establishment of proper standards of attainment, and the conduct of examinations, nothing of a similar nature exists in the parallel, and not less important, sphere—at least as far as engineering is concerned—of practical training.

In its consideration of sandwich courses submitted for approval for the new Diploma in Technology, the National Council for Technological Awards insists on a minimum aggregate period of two years' practical training, and it has appealed to the firms with which this training is being undertaken to assist in ensuring that the content of the College and industrial parts of the courses are closely co-ordinated. Yet against the care it has taken in assessing College facilities and staff and the structure and curricula of the College courses, and its requirement for the appointment of external examiners, it has not been able to make a similar assessment, nor to impose any real safeguards on the nature and quality, of the associated industrial training. No doubt the Colleges will do what they can in these respects and they may be able to derive strength through the recent decision of the Ministry of Education authorising Local Education Authorities to provide maintenance grants, and to pay the fees, for sandwich course students while at College where this assistance is not forthcoming from their employers. It would seem to me desirable that in making these awards the Local Education Authorities should satisfy themselves that the training which the students are undergoing during the works periods is of a standard acceptable to the Colleges, and through them to the National Council.

This requirement would not be easy to impose at a time when the Colleges have a difficult enough task in ensuring that their approved courses are adequately supplied with students from industry of the right academic quality, and when they are short of staff. But I hope the time will come when they will feel able to question the acceptability as Dip. Tech. students of boys submitted to them by firms with whose training arrangements they are not satisfied. Unfortunately, this could inflict hardship on some able boys—but no more so perhaps than is involved in the non-acceptance of a particularly convenient College for the purpose of the Diploma in Technology award. Just as in the latter event the boys will go to another, approved, College, so they, or their schoolmasters, will soon discover which firms are to be avoided. The pity is that these firms may not greatly care, notwithstanding that they are probably the ones most in need of able recruits; and the anxiety that

as we pass through the 'bulge' of young people seeking employment during the next few years the training capacity of the remainder may be too small.

Though the scale of the industrial facilities for the satisfactory training of graduate and student apprentices is inadequate, and the measure of supervision which can be exercised over them from outside is small, there are at least signs of machinery within some of the professional Institutions as well as the Hives Council by which appropriate pressures may be applied. But even these do not yet exist in respect of the nationally no less important spheres of technician and craft training. This is perhaps understandable in the case of technicians since their emergence as a category of industrial personnel for which distinctive schemes of technical education and practical training are required is of much more recent occurrence than for craftsmen. Of our inadequacies on the craft side of industry on the other hand one can only say that they are potentially ruinous. We have been proud to claim that British workmanship is the best in the world, and indeed there *has been* much truth in this claim. We shall be wise to recognise, however, that the justification for the claim is wearing precariously thin, for it has been founded not on a high standard of craft training so much as on the past entry into craft employment of very many boys of high natural ability—boys who became excellent craftsmen in spite of the deficiencies of their prior training. From now on an increasing proportion of the equivalent of these boys will remain at school beyond the age at which, under present circumstances, they can be recruited by industry as craft apprentices, if indeed they would then wish to enter it in this category. We must, therefore, expect a lowering in the quality of the entrants to craft apprenticeship at a time when what is wanted is a raising, not a lowering, of the standard of craft skill. This need cause no anxiety granted ample, and good, training facilities. The problem is how to get them.

The highly unsatisfactory present state of craft training in the bulk of British industry has been discussed in a most searching and constructive manner in a recently published book by Lady Gertrude Williams entitled *Recruitment to Skilled Trades*. Her excellent analysis of the fundamental changes which might be introduced raises a whole range of questions. What duration of training, suitably organised and conducted, is really necessary for the different crafts, as against the universal five-year period from 16 to 21 years of age for all crafts? Should not the age of entry to craft apprenticeship be raised to permit the entry of boys who have pursued general education at school beyond the age of 16 years? Should not the early stages of training be carried out in training centres on a full-time basis where continued general education could also be catered for? Can such centres be organised co-operatively by groups of firms on an adequate, nation-wide, scale, or must they be provided by the Local Authorities? How should the numbers of recruits for the various crafts be assessed? What tests of competence should be introduced, particularly at the end of apprenticeship, and what should be the significance of the results of these tests in respect of subsequent employment? How should training be financed—in particular how should individual firms be assisted to bear the cost of running their own training schemes?

The extraction of satisfactory answers to these and other related questions, and their implementation on a national scale as a matter of urgency, will involve considerable changes in attitude among both Employers and Trade Unions, an extensive building programme and the solution of a difficult staffing problem. They provide infinite scope for discussion and I might have devoted a large part of this address to an expression of my opinion about them. But I doubt whether any individual opinion will make much impact in the continued absence of a national body with executive powers and the financial means to exercise these powers effectively. The report on *Training for Skill* published in April last by a Sub-Committee of the National Joint Advisory Council under the Chairmanship of

Mr. Robert Carr went no further than to envisage a purely advisory National Apprenticeship Council intended to 'keep representative organisations of employers and workers at national level informed of progress' and to 'collect and disseminate information about these aspects of training which are common to more than one industry'. A 1944 report of a Conference of Representatives nominated by the Manchester and District Committees of the Engineering and Allied Employers National Federation, the Amalgamated Engineering Union, the major Engineering Institutions and the Regional Advisory Council for Technical and other forms of Further Education, went much nearer the mark.

This report proposed, *in 1944*, the establishment of a National Engineering Apprentice Council as a central *directing* body, and of associated Regional Councils, each to be composed of representatives of organised Employers, Trade Unions, Professional Institutions, Education Authorities and the interested Government Departments, and it defined the responsibilities of the Regional Councils for the organisation and administration of training schemes in their respective areas as follows:

(1) to estimate the future needs of the industry in its area and to adjust the intake of apprentices to training schemes accordingly;
(2) to advise upon the establishment, equipment, staffing and maintenance of all Apprentice Training Centres (for the initial period of training) to meet the needs of the region;
(3) to co-operate with employers in the placing of apprentices who have completed their courses in the Apprentice Training Centres;
(4) to approve facilities for, and maintain contact with, the (further) training of apprentices during the period of their apprenticeship spent in industry;
(5) to interest itself in the development of the non-vocational interests of the apprentices;
(6) to issue certificates of competency to apprentices on completion of their term of apprenticeship.

The report then states:

The Conference takes the view that the initial part of the proposed training scheme should be financed by a National Fund provided jointly by the State, the Local Education Authorities, Employers of engineering craftsmen, and the Trade Unions concerned, on a basis to be determined after a detailed investigation. The financial arrangements under the scheme should cover not only any necessary expenditure on the provision and maintenance of Apprentice Training Centres, but also the maintenance of the Apprentices during the first two years of training under the Scheme.

For the final period of training (i.e. that during which the apprentice is apprenticed to an individual firm) the Conference considers that his wages and the cost of his training should be met by the firm concerned, but that any fees chargeable on account of his attendance at technical classes under the scheme should be met from the National Fund.

The difficulties which would be associated with the implementation of such proposals as these would be considerable. But the issues are of great urgency, and the difficulties must be faced and solved.

The Special Committee on the Supply and Training of Technical Teachers of which I was privileged to be Chairman concluded the outline of its report as follows:

The task which our enquiries have disclosed is a difficult one, but it is not impossible to accomplish. It is within the country's power to ensure that its system of technical education, both in its teaching staff and in its buildings, is made adequate to the demands which will fall upon it in the next decade. Success will depend upon the energy, discipline and co-operation of the partners concerned in this great enterprise.

I cannot close with better words, taken in the wider context of my address. I would only add that the more effectively industry fulfils its dual responsibility as

one of the partners, the fuller will be the benefit which accrues to it—employer and employee alike—and through it to the nation.

Appendix I

Distribution of manufacturing concerns expressed in terms of the number of employees December, 1955*

Total number of firms—206,300 approx.
Total number of employees—9 million approx.

Number of employees	Less than 10	11–24	25–99	100–499	500–999	1,000–1,999	2,000–4,999	More than 5,000
Number of firms	150,000 approx.	15,509	25,830	12,178	1,666	723	328	79

* Ministry of Labour Gazette, Dec. 1956.

Appendix II

*Growth since 1938 of the numbers of school boys and girls passing the Advanced Level G.C.E. examinations in Physics, Chemistry and Mathematics**

Year	Physics			Chemistry			Mathematics†		
	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls	Total
H.S.C.									
1938	Not available		3,153	Not available		3,035	Not available		3,224
1948			8,589			7,196			7,508
G.C.E.									
1955	11,429	1,642	13,071	9,624	1,768	11,392	9,798	1,214	11,012
1956	12,677	1,705	14,382	10,136	1,828	11,964	10,938	1,400	12,338
1957	13,639	1,791	15,430	10,829	2,150	12,979	12,325	1,739	14,064

* Supplied by the Ministry of Education.
† Excludes Applied Mathematics and Mechanics.

Appendix III

Anticipated school leavers at different ages in England and Wales The peak year for each age group is shown in brackets*

Children who in the year they leave school will reach the age of	1956	1962	1963	1964	1965	Percentage increase of peak year over 1956
15	333,000	(494,000)	430,000	401,000	377,000	48
16	68,000	109,000	(121,000)	109,000	105,000	78
17	34,000	49,000	62,000	(70,000)	64,000	106
18 and above	31,600	49,000	49,400	55,800	(66,200)	109
Totals	466,600	(701,000)	662,400	635,800	612,200	50

* *Training for Skill*, p. 1. Report of the Carr Committee. March 1958.

The numbers of boys and girls leaving school at ages above fifteen in Scotland are also expected to rise, though less rapidly than in England and Wales.

APPENDIX IV

Growth since 1938 *of students qualifying for Ordinary and Higher National Certificates**

Subject	1938		1957	
	O.N.C.	H.N.C.	O.N.C.	H.N.C.
Mechanical Engineering	1,449	502	8,063	3,590
Production Engineering	—	—	—	521
Electrical Engineering	917	379	3,595	2,048
Civil Engineering	—	—	—	270
Building	544	137	1,436	758
Chemistry	151	62	1,263	755
Applied Chemistry	—	—	20	32
Metallurgy	—	—	288	222
Applied Physics	—	—	254	112
Commerce	132	—	381	9
Naval Architecture	16	—	75	76
Textiles	104	57	132	94
Mining	—	—	669	181
Mining Surveying	—	—	—	92
Chemical Engineering	—	—	—	36
Totals	3,313	1,137	16,176	8,796

* Supplied by the Ministry of Education.

APPENDIX V

Composition of part-time teaching staff in Technical Colleges. Session 1955–56*

(i) Full-time teachers whose college teaching forms part of their full-time service (e.g. in secondary technical schools)	1,359	3%
(ii) Full-time teachers whose college teaching is an extra to their full-time service (e.g. evening teachers)	10,310	24%
(iii) Those 'engaged in the professions, industry or commerce, teaching their own or cognate subjects'	26,508	60%
(iv) Others	5,801	13%
Total	43,978	100%

* *The Supply and Training of Teachers for Technical Colleges*, p. 50. H.M. Stationery Office. May 1957.

DIGESTION IN THE RUMINANT[1]

ADDRESS BY DR. D. P. CUTHBERTSON, C.B.E.

PRESIDENT OF SECTION M

Introduction

The sub-order *Ruminantia* includes the chevrotain, deer, giraffe, okapi, pronghuck, ox, yak, goat and sheep, but other mammals have multiple chambered stomachs, e.g. bats and some of the whales. Australian workers have demonstrated that the wallaby (*Setonix brachyurus*) shows a degree of convergent evolution in that it has a similar stomach system and blood sugar level (Moir, Somers and Waring, 1956). But for the purposes of this address we are concerned with our domestic cattle, sheep and goats, for these ruminants owe their place in our agricultural economy to their peculiar digestive system.

The outstanding difference between the ruminant and other herbivores is that the former pre-ferment their food through microbial action in the first two compartments of the complex stomach—for convenience termed the reticulo-rumen sac—before digestion by enzymes secreted by the digestive glands begins in the abomasum or fourth stomach. Fermentation of cellulose through microbial activity occurs in the sac and to a limited extent in the caecum and colon of herbivores. Digestion is not as efficient in these parts as in the rumen, possibly because the rate of passage is too rapid to permit effective bacterial action.

The reticulo-rumen sac thus forms a very special kind of fermentation vat in which bacteria and protozoa are established under highly reducing conditions when the animal is quite young and the rumen is relatively small. Removal of much of the cellulose, and growth of micro-organisms both precede the normal alimentary digestion and absorption. This enables the products of microbial synthesis and the residual food nutrients enclosed within the plant cell walls to become readily available for absorption.

In grass itself, and especially in root crops, other fermentable carbohydrates exceed cellulose in amount and being more soluble and less structurally and mechanically resistant than cellulose, are more readily fermentable by a much wider variety of micro-organisms but not necessarily to the same ratio of desired products. Whether the products of such fermentations are more or less suitable for the metabolic processes of the ruminant than the carbohydrates themselves is a matter for speculation to which there is still no answer. My late colleague Frank Baker (1946) described the reticulo-rumen process as a functional field in which all the activities are organised for utilisation of microbial products and all factors therein make for the attainment of a steady state.

The reticulo-rumen sac is normally never empty and this is probably why the gastric and intestinal digestive secretions are continuous (Hill, 1955). The activity of the salivary glands, the process of rumination involving re-mastication, the absorption of the products of microbial activity and the interchange of ions between rumen contents and blood all contribute to the continuity of the microbial action during the whole life of the animal and produce conditions which one can hardly hope to reproduce *in vitro*. Nevertheless, by a study of the micro-organisms—the bacteria and protozoa, the latter chiefly ciliates—at the site of their attack on food particles, and in isolation in culture, and by a study of the soluble products of

[1] Address delivered on Thursday, August 28, 1958, at the Glasgow Meeting of the British Association.

fermentation and their absorption, much information about this microcosm has been obtained. (For earlier reviews see Hungate (1950*a*, *b*), Doetsch and Robinson (1953), Huffman (1953), Underwood and Moir (1953), Owen (1954), Chalmers and Synge (1954*b*), Edwards (1955), Phillipson and Cuthbertson (1956) and Maynard (1957).

Early Development

The four chambers of the ruminant stomach are well defined at birth, the abomasum being proportionally the largest compartment. The rumen and reticulum are lined by an epithelium of oesophageal type, prolonged into low villous processes—in the case of the reticulum a network of ridges. The omasum also has a wall of oesophageal structure, folded into long and very muscular ridges, by which the mixture of digesta is further broken up. The abomasum is the site of gastric secretion and apart from the large, curiously spiral colon this and the rest of the alimentary tract is like that of the rat and man.

Normally, when the young ruminant sucks its mother the milk passes directly to the abomasum owing to reflex closure of the oesophageal groove. This direct passage of milk to the acid environment of the abomasum where peptic digestion occurs is a means of avoiding undesirable bacterial degradation of the milk constituents. By 48 hours after birth the flow of gastric juice seems well established and so further bacterial activity in the abomasum is unlikely. Up to this time the passive transfer of colostral proteins concerned with immunity can take place. Salivation in the newborn animal is undeveloped.

By 6 weeks of age the rumen contents appear normal in so far as smell, colour and consistency are concerned. The adult bacterial association is not built up at once and before rumination is established the rumen population is apparently quite different from that found in the adult.

Lower volatile fatty acids result from the fermentative processes in the reticulo-rumen sac—a fact known since Tappeiner's interesting study of 1884, which was confirmed by Henneberg and Stohmann in 1885 but for long forgotten until Phillipson's work in 1942—and a comparison of these at different ages indicates that during the first 3 months of life the proportions differ from the general adult pattern in that more propionic and less acetic are present in the early stages. In adult sheep and cattle the proportions, while varying with the food, usually are in the following order of descending magnitude: acetic, propionic, *n*-butyric, *iso*-butyric, and isomers of valeric acid. Traces of formic acid are reported and small traces of succinic. Gray, Pilgrim, Rodda and Weller (1951, 1952) suggested that branch-chain isomers were derived from the breakdown of proteins and el-Shazly (1952*a*) noted that they increased when ammonia was being formed.

The preponderance of lactobacilli in the reticulo-rumen sac of the young calf (Mann and Oxford, 1954, 1955; Huhtanen, Saunders and Gall, 1951) predisposes to the idea that lactic acid is also present in quantity and a large Gram-negative coccus is found in considerable and relatively constant numbers in young calves but disappears from the rumen as the animal grows older (probably at about 4 months old) and is not found, or is found only in small and irregular numbers in adult cattle and sheep (Hobson, Mann and Oxford, 1958, and to be published). This organism will ferment lactate to fatty acids including propionic acid (Elsden, Volcani, Gilchrist and Lewis, 1956; and Hobson, Mann and Oxford, *loc cit.*) and its presence in the rumen of young calves may to some extent account for the occurrence of a larger proportion of propionic acid in these animals than in older ones. Since lactic acid is a common product of the fermentation of sugars by a large number of rumen bacteria and since it is not usually found in any large concentration in the adult animal and is not all absorbed, it must be fermented by rumen bacteria other than this large coccus. Organisms capable of doing this and also other fermentations have been found (Johns, 1951; Gutierrez, 1953; Hobson

and Mann, unpublished). Lactic acid may certainly be present in appreciable concentrations when starch or starch-rich rations are fed.

Since the young ruminant can store body nitrogen and grow when given a protein known to be of fair nutritive value for the rat and yet will lose body nitrogen and weight when given an inadequate protein, it obviously is dependent on its diet for at least part of its amino acid supply (Blaxter and Wood, 1952).

Protozoa do not appear in the reticulo-rumen until calves are about 1 to 2 months old (Lengemann and Allen, 1955) and contact with normal animals harbouring these appears to be necessary as no one has yet proved the existence of free living forms of these rumen organisms. The origin of the numerous bacteria in the rumen is not known, but a large number are no doubt ingested with the feed. The use of cud inoculation hastens the development of the adult flora in the rumen and has been advocated by some, and while there is no doubt it establishes a protozoal population more rapidly than otherwise, no clear-cut overall benefit to the animal has yet been established (Pounden and Hibbs and their associates, 1948*a*, *b*; 1949; 1950*a*, *b*). The feeding of good quality hay has also been recommended as this encourages the early establishment of the adult bacterial population as far as can be judged by microscopic appearance.

As lambs start to nibble grass at 14 to 18 days of age and by 3 weeks are eating considerable quantities of hay (Munro, 1956), there seems little reason to delay the treatment of the young ruminant as an adult animal longer than about 6 weeks in the case of calves and 3 weeks in the case of lambs from the point of view of feeding, providing adequate milk is given.

Under natural conditions the young ruminant acquires its ciliate protozoan population by direct mouth contact from animals possessing the fauna. It was Becker and Hsiung (1929) who finally proved experimentally that only direct contact allowed transfer of these organisms and that no resistant forms were to be found in the faeces or in the diet. Whether or not the organisms transferred become established is, however, determined by the conditions within the rumen and consequently by the diet. It was because a hay diet provides a suitable medium and consequently leads to a rise in ciliate numbers that spores were earlier believed to exist in the hay.

The environment imposed by the ration is of importance when studying the fauna of young ruminants. Where conditions within the rumen are acid as, for example, when lactobacilli are numerous as a result of feeding high starch rations, the ciliates rapidly die off. This appears to be the reason why under normal management routine, ciliates do not become established in calves until they are 8 weeks old or more. On the other hand, frequent inoculation of calves under early weaning conditions (Preston, 1956) in conjunction with examination of the microfauna and flora have shown that provided roughage is available, ciliates will develop in animals under 2 weeks of age while large quantities of milk are still being fed. However, when this diet is supplemented with a meal ration lactobacilli appear in large numbers, the fauna again disappears and the overall rumen microbial picture reverts to what is generally regarded as typical of the young animal. Work is in progress at the Rowett Institute to determine if the dilution of the meal with roughage in an endeavour to establish and *maintain* an adult type microflora and fauna at an early stage, would be valuable in early weaning. In addition, because of the comparative simplicity of preventing ciliate contamination by isolation, work is being carried out with faunated and unfaunated twin calves. Although Pounden and Hibbs (1950*b*) found little difference between calf twins with and without ciliates Dr. Margaret Eadie at the Rowett Institute is trying to determine whether the fauna proves beneficial when the animal is subjected to stress such as a sudden change of diet.

Just as the numbers of ciliates vary with diet, so also do the relative numbers of

species, thus indicating considerable variation in the properties of these organisms. It has been possible to defaunate an animal partially and then isolate separate species in a single sheep (Eadie and Oxford, 1957). Thus the properties of chosen organisms may be examined under natural conditions and in pure washed suspensions without the difficulties and possible errors of attempting to maintain and study pure cultures. The biochemical properties of certain of these ciliates are being examined at the Rowett Institute by Dr. B. H. Howard (Howard, 1957).

Saliva and Ruminant Digestion

Of the salivary glands, the parotids have been most extensively studied as they supply a large part of the water and alkali required to preserve the fluid nature and hydrogen ion concentration of the digesta in the rumen. Preliminary experiments on calves by Dr. R. Kay (unpublished) have shown that the parotids secrete little saliva in the first month of life, at which age the rumen is small and relatively empty. This may account for difficulties encountered in the early weaning of calves. Thereafter the weight of the glands increases approximately in proportion to body weight. They secrete rapidly during feeding and rumination, and at a slower rate when the animal is idling. In adult sheep they secrete 1–4 litres each per day (Denton, 1957; Kay, unpublished).

The rate of secretion of the parotids is most rapid when rough fodder (hay, oats, etc.) is being eaten, and is much less affected by chewing grass or roots (Colin, 1886; Scheunert and Trautmann, 1929 (sheep); Balch, unpublished (cattle).

Reflex salivary secretion may be concerned in the etiology of bloat; increased secretion might either mitigate the condition by making the rumen contents more fluid or worsen it by adding carbon dioxide bubbles and mucus to the froth.

Submaxillary and sublingual saliva is secreted in much smaller quantities and more weakly buffered than parotid saliva (Colin, 1886). Two types of gland—the inferior molars and small mucous glands—appear to contribute to the residual secretion (Kay, unpublished).

The Breakdown of Protein in the Rumen and its Significance in the Nutrition of Ruminants

In my presidential address to Section I in 1953 I outlined the origins of our interest in this aspect of digestion and utilisation of proteins and other nitrogenous compounds by ruminant animals. The results of these observations and all later experiments of a comparable nature at the Rowett Institute have tended to show that the rate of evolution of ammonia in the rumen first noted by McDonald (1948), is nutritionally important and must be taken into account when assessing the value to the ruminant animal of various sources of nitrogen. My colleague Dr. R. L. M. Synge has guided our researches in this field since these early observations. The work of Chalmers and Synge (1954*b*) and that of their colleagues, el-Shazly (1952*a*, *b*), Annison (1954), and Warner (1956*a*, *b*) can be summed up as indicating that it is reasonable to suppose that dietary protein is subjected in the rumen to fairly vigorous hydrolysis conditions. The low-molecular hydrolytic products, including amino acids, may then be assimilated by micro-organisms and incorporated into their body proteins or, if amino acids accumulate in excessive amounts, the micro-organisms acquire the power to ferment them deaminatively to ammonia, fatty acids and carbon dioxide. It is also known that urea and ammonia can be utilised by rumen micro-organisms for elaboration of their own body proteins. We are now only beginning to recognise some of the micro-organisms concerned in this.

It would appear that proteolysis is probably carried out by a large variety of the rumen bacteria, some of which are facultatively anaerobic protein-splitting types derived probably from the foodstuffs (Blackburn and Hobson, unpublished obser-

vations). Bacteria have been isolated from the rumen, which have been shown to utilise the nitrogen of ammonia for synthesis of bacterial protein (Wasserman, Seeley and Loosli, 1953).

At the Hannah Institute Pearson and Smith (1943) had earlier shown *in vitro* that the value of a protein to the ruminant might depend on the extent to which it was degraded in the reticulo-rumen sac. On the more practical side, observations by Chalmers and Synge (1954*a*) indicated that hardening casein by heating and then grinding finely, could protect the protein from vigorous deamination and yet allow it to be digested. Such material, which was of lower nutritive value to the rat than the material from which it was prepared, was nevertheless of increased value to the ruminant. Experiments using herring meals, groundnut meals, etc., suggested that differences in processing might have significant effect on the nutritive quality of the product for ruminants. Some of the treatments that gave the best meals for ruminants might also be advantageous from the technical side as concerns processing for oil extraction.

In a detailed investigation on sheep it became obvious that the extent of ammonia evolution from different proteins in the reticulo-rumen sac only becomes a decisive factor in protein nutrition when the protein content of the diet is fairly high. With lambs receiving a typical fattening ration, it made no difference whether the protein supplement was casein or herring meal. Nevertheless, on rations sub-optimal in quantity but having protein content similar to that of dairy rations, lambs grew strikingly better when the main source of protein was herring meal (which is not so readily attacked), rather than casein, and the differences in wool production were even more striking. It would seem that these differences between protein sources may be important when animals have to receive supplements to a low diet also low in protein, as under Australian drought conditions, where wool production is an important consideration. However, under British conditions, ammonia formation seems to Synge and his group to be more likely to be an important factor in protein nutrition in dairy diets.

In lactating goats it has been shown conclusively by Chalmers and Synge (1957) that at all stages differences in the direction to be expected can be gauged from differences in ammonia evolution in the rumen. Indeed, preliminary tests on the time course of evolution of ammonia in the rumen are generally a useful exploratory measure of studying the utilisation of protein in the ruminant. But in experiments comparing copra and undenatured groundnut meal in goats the evolution from the copra was very low compared with the groundnut meal, but in a nitrogen-balance experiment on the goats no difference in nitrogen retention or digestibility was found. In rats copra did not support growth. The ammonia curves could be depressed by addition of starch or fat.

Further nitrogen-balance tests on goats fed undenatured groundnut meal, toasted groundnut meal, and low-temperature groundnut meal showed a correlation between ruminal ammonia production and nitrogen retention figures. Low-temperature dried meal gave the highest ammonia production and the poorest nitrogen retention figures. In a growth experiment test on rats, the toasted meal caused a decrease in growth of 25 per cent compared with the undenatured groundnut meal, whereas this toasted material gave a low ammonia production in ruminants and the best nitrogen retention figure of the groundnut meals tested. Thus, while heat-treatment damages the protein for the non-ruminant, the resulting slower microbial attack in the ruminant permits the elaboration of microbial protein which is useful to the animal. Similar experiments on differently processed herring meals showed the same trend.

Results on grass to date indicate that a mixture of a high-protein, low total water-soluble carbohydrate (TSC) grass with a low-protein, high TSC grass gave better nitrogen retention than did either the high- or the low-protein grass. Dried

grass of 20 per cent protein was fed with molasses, starch or hay to give a 14 per cent protein ration and these were given in equal bulk and equal nitrogen content. The rations with the carbohydrate in the form of molasses or starch gave a nitrogen retention of 1 to 2 g. per day more than the hay. If the nitrogen of urea added to a low-protein ration is to be of use to the animal then soluble carbohydrate must be fed in adequate amounts.

Synthesis of Microbial Proteins from Ammonia (Urea) and Amino Acids

Ever since the last quarter of the nineteenth century the possibility that rumen micro-organisms might synthesise their body protein from simple nitrogenous sources in the diet, thus supplementing the protein supply of the host or at least saving dietary protein from attack, was entertained. Most work has been done on urea and there is substantial agreement that much carbohydrate must be present if there is to be effective utilisation of the urea nitrogen. Simple sugars are fermented too rapidly and the digestion of cellulose and related polysaccharides does not seem to provide the necessary conditions, but on the other hand starch is most useful. It seems likely that urea and other forms of non-protein nitrogen will not be effectively used when other proteins in the diet are undergoing substantial degradation to ammonia; thus, the nature of the protein fed may also affect the utilisation of urea. We know very little as yet about the extent to which the different rumen micro-organisms use extracellular ammonia, free amino acids or peptides.

Loosli, Williams, Thomas, Ferris and Maynard (1949) have studied the synthesis of amino acids in the rumen of sheep and goats by feeding a synthetic diet in which urea supplied almost all the nitrogen and found considerable synthesis of all the amino acids analysed, the increase in lysine being particularly high. The work of Bouchaert and Oyaert (1952) and of Duncan, Agrawala, Huffman and Luecke (1953) shows that with urea as practically the sole source of nitrogen the ruminant can readily synthesise the amino acids that are essential in the diet of the rat.

Lactobacillus bifidus isolated from the rumen of sheep fed a high flaked maize ration use ammonia nitrogen *in vitro* in the presence of cysteine alone or cysteine and peptone (Blackburn, Dobson and Phillipson, unpublished). Absorption of ammonia from the rumen is favoured by the concentration gradient, for the content of ammonia in the peripheral blood is very low unless protein feeding is heavy and even then it is small compared to the concentration in the rumen. The only way to reduce ammonia concentration in the rumen is to feed starch supplements which have the effect of stimulating ammonia utilisation.

Digestion of Carbohydrates

A review of the literature dating from the experiments of Kellner (1913) indicates that before about 30 days of age the calf is unable to use starch modified either by cooking or by enzyme action. Scouring, and in many instances death, results when calves are fed diets containing modified starches. The reason for the failure to utilise starch or the products formed from it by the action of diastase, and so to grow, is not certain, but it is presumed to be absence of development of the appropriate enzymes. Of monosaccharides only glucose and lactose appear to be well utilised but the total amount of sugar that may be fed is limited probably by the capacity of the absorptive processes and coincident exposure to *E. coli* infection. Starch, dextrin, and maltose are not utilised by young calves (Dollar, 1958), and Dollar and Porter (1957) have shown that maltase is not present in mucosal extracts of the young calf.

When cellulose is broken down by the micro-organisms to soluble sugars, these, like other soluble sugars in the diet, are fermented to give short-chain fatty acids. In grass itself, and especially in root crops, other fermentable carbohydrates may exceed cellulose in amount, and are more readily fermented by a much wider

variety of organisms. Little is known about the early stage of cellulose cleavage into soluble products. The nitrogen need of the cellulose-splitting organisms may lie behind the need for an adequate supply of protein or other nitrogenous compounds for satisfactory digestion of cellulose and other roughage polysaccharides. The depressing effect of starch or molasses on cellulose digestion may be due to competition for available nutrients by the more rapidly-growing organisms that ferment available carbohydrate. The fermentation products of cellulose include short-chain fatty acids, carbon dioxide and hydrogen. Carbon dioxide forms some 60 per cent of the gas in the reticulo-rumen and is derived from both the food and the saliva. Methane arises by reduction of the carbon dioxide by bacteria. Marston (1948) was able to give a balance-sheet which indicated the recovery of most of the carbon of cellulose arising from its *in vitro* fermentation in the form of propionic and acetic acids. Smaller quantities were found as butyric acid, methane, carbon dioxide and other minor products. The proportion of energy incorporated in the bacteria themselves is approximately 6 per cent of the total used.

Seaweed-eating ruminants can live quite successfully with practically no cellulose. The carbohydrate content of their diet is made up of substances such as mannitol, laminarin, fucoidin, xylan and alginic acid, and would be of little value to the non-ruminant. The protozoa of these sheep in North Ronaldshay, Orkney, which feed practically exclusively on seaweed is similar to that found elsewhere (Eadie, 1957). Though there was evidence of less storage carbohydrate material within certain of the protozoa most noted for their rapid fermentation of soluble carbohydrates (Oxford, 1951), these organisms were present in considerable numbers. They also proved to be able to use laminarin *in vitro*, a property not noted when studying organisms from sheep on a normal diet. This property was of particular interest since the sheep appeared to have a particular preference for the seaweed *Laminaria*.

Pentosans, on which very little work has been done, form some 18 per cent of the dry matter of grass and hay, and Howard (1955) has shown that mixed rumen bacteria will split these with the formation of simpler sugars and then ferment the products to volatile fatty acids.

Bacteria capable of fermenting practically all fodder carbohydrates have been isolated from the rumen and are present in sufficient concentration to cause extensive degradation of these materials there. Rumen ciliate protozoa compete with bacteria for some of these fodder carbohydrates. It is appropriate to place bacteria and protozoa on an equal footing in this respect.

Glucose, sucrose, inulin and soluble levans are fermented by *Streptococcus bovis* —which is also the main starch fermenter—and by the protozoa. From starch/lignin ratios Weller and Gray (1954) have shown that very little starch reaches the abomasum even when large amounts are fed. One possible benefit of feeding the antibiotics such as aureomycin may be through suppression of lactobacilli in the abomasum.

What escapes microbial digestion is presumably in greater or less degree digested by the carbohydrate-splitting enzymes of the alimentary secretions.

The main end products of the fermentation of these carbohydrates by both bacteria and protozoa are acetic, propionic and butyric acids. Where a large quantity of starch, e.g. flaked maize, is fed, a lactic acid fermentation is set up and this is associated with a decreased production of acetic acid and an increased proportion of propionic acid (Phillipson, 1952). This is the result of the lactic being further fermented to acetic and propionic acids. Associated with hydrolysis of polysaccharide by micro-organisms is storage of starch-type polysaccharide within the cytoplasm. In the case of the protozoa this is an amylopectin (Hobson and Mann, 1955; Doetsch, Howard, Mann and Oxford, 1957).

While it is generally agreed that little or no digestion of lignin takes place (Gray, 1947), the observations of Pigden and Stone (1952) need extension, for they found

that although the lignin in dicotyledenous forages is relatively unchanged by passage through the ruminant, the lignin in monocotyledenous hays is partially digested.

Lipids

From the quantitative standpoint lipids do not form an important constituent of the diet of ruminants. The dry matter of cocksfoot, for example, contains only from 2 to 4 per cent of long-chain fatty acids—predominantly in unsaturated form. As the body fat of ruminants is largely saturated and not susceptible to change, the suspicion that rumen micro-organisms may hydrogenate unsaturated fatty acids has now been demonstrated by Reiser (1951) and Shorland, Weenick and Johns (1955). Rumen contents effectively hydrogenate linoleic and oleic acids. As the sheep must rely on dietary sources of essential fatty acids (Cunningham and Loosli, 1954), it is important that hydrogenation of dietary essential acids should not proceed to completion.

Nutritive Properties of Rumen Micro-organisms

It is reasonable to believe that, metabolically, ruminants have specific needs for amino acids similar to those of rats and man, but they do not require to have their specific needs in their feed because the rumen micro-organisms can build their body protein from various non-protein nitrogen compounds other than amino acids of the feed protein.

Since Müller's work in 1906 it has been known that the experimental feeding of rumen micro-organisms to simple-stomached animals has consistently shown them to be of good nutritive value. In chicks, Usuelli and Fiorini (1938) found that the protozoal fraction gave rather better growth than the bacterial fraction. Following this up, the observations of McNaught, Owen, Henry and Kon (1954) showed the somewhat higher value of the protein of the protozoa to rats was due to their higher digestibility.

The first signs of disruption of the bodies of the bacteria take place in the omasum. It is assumed that they subsequently become available to attack by the alimentary enzymes. As the protozoa apparently require pre-formed protein or amino acids for growth they do not contribute to the overall protein synthesis. Both bacteria and protozoa are thus valuable sources of protein for these higher animals. It is estimated that about 50 per cent of the nitrogen of rumen contents in hay-fed sheep is microbial protein nitrogen. The seasonal variations in the carbohydrate of grass are probably of importance in the utilisation of the protein and non-protein nitrogen. That there are no 'essential fatty acids' in rumen bacteria has been demonstrated by Cunningham and Loosli (1954) and Garton and Oxford (1955).

Vitamin Synthesis

Studies on goats (British Saanen) indicate that while carotene as such is absorbed only in traces and so cannot be determined in blood or milk, at parturition there is a mobilisation of carotene so that the colostrum contains much carotene and abundant vitamin A (Chanda and Owen, 1952). It has been shown that while there is no intestinal synthesis of carotene in the goat, nevertheless due to the continual presence of carotene in the blood of the cow (but not of the goat) there is a small but measurable residual defaecation of carotene by this ruminant. Dr. E. C. Owen (private communication) does not believe that the microbiological synthesis of any fat-soluble vitamin by the ruminant (other than vitamin K) has yet been unequivocally demonstrated.

While ruminants have apparently similar metabolic needs for the various B vitamins as do other species, they have little, if any, dietary requirements for them because of the synthesis which occurs in their digestive tracts, primarily in the

rumen (Kon and Porter, 1954). The synthesis of the members of the B group appears to develop *pari passu*. Vitamin B_{12} is a special case because its synthesis is conditioned by the cobalt content of the crops and these may be deficient in certain areas.

The synthesis of the B group of vitamins (thiamin, riboflavin, pyridoxin, niacin, pantothenic acid, biotin, folic acid and vitamin B_{12}) by bacteria in the reticulo-rumen is associated with the development of the digestive tract and occurs quite early in young calves receiving solid food. There is also evidence of their synthesis in the large intestine (see review by Kon and Porter, 1954). Usuelli (1949) showed that the rumen protozoa do not readily synthesise vitamins of the B group.

There is a general constancy of the levels of B vitamins in milk under different dietary treatments (Kon and Henry, 1949; 1951).

Marston and Smith (1952) have pointed out that the requirement of the ruminant for vitamin B_{12} is high compared with the non-ruminant and pining, as is well known, occurs in the absence of sufficient cobalt in the diet.

Of the cobalt-containing compounds of the vitamin B_{12} family in the rumen, vitamin B_{12} itself, cyanocobalamin, represents only a small part of the total content. Factor A, a purine analogue, is the dominant member in rumen contents but no organisms have yet been found that form appreciable amounts of it (Ford, Kon and Porter, 1951; Porter, 1953, 1955; and Kon, 1955). We know that cyanocobalamin is used by the ruminant and the related analogues could potentially be of importance to the rumen micro-organisms concerned in the dissimilation of feed to lower fatty acids (Dollar, 1958).

Ascorbic acid is not synthesised by micro-organisms but is made in the cow's own tissues.

Minerals

Mention has been made of the important rôle of cobalt in vitamin B_{12} synthesis. Copper is mainly associated with the insoluble matter of the plant cells (Mills, 1958) and its deficiency is associated with disorders of many systems. McNaught, Owen and Smith (1950) found that protein synthesis ceased *in vitro* when the copper fell to 12 p.p.m., and cobalt at the low level of 10 p.p.m. also reduced synthesis. Burroughs, Latona, De Paul, Gerlaugh and Bethke (1951) found that cellulose digestion could be encouraged by addition of minerals and they concluded that the decrease in roughage digestion in the presence of maize starch could be explained on the basis of the mineral requirements of rumen bacteria.

Dr. Rook will describe the observations which he and Dr. Head have made on the level of ammonia production in the rumen and magnesium absorption.

Feeding of Antibiotics

It is established that low-level feeding of the tetracyclines (20–100 mg. daily) to young calves increases their rate of live weight gain, food consumption and efficiency of food conversion, and decreases the incidence and severity of scours (review by Lassiter, 1955). The inclusion of 75 mg. of oxytetracycline in predominantly concentrate rations fed to adult beef cattle in the U.S.A. has also been considered to increase significantly growth and food efficiency (Pfizer, 1957). It has been reported that chlortetracycline has been fed to milking cows at 0·1 mg. per lb. of body weight without it being detected in the milk and with benefit in some cases of respiratory and other infections, and also in foot rot.

With early-weaned calves the major part of the increased growth rate of those fed antibiotics can be accounted for by their greater consumption of food relative to control animals (Preston, McLeod and Dinda, unpublished).

Rumen pH is higher in calves fed an antibiotic than in controls (Mann, Masson and Oxford, 1954; Preston, McLeod and Dinda, 1958) and this is due to a lower

concentration of volatile fatty acids in the rumen (Preston *et al.*, 1958; Johns, Mangan and Reid, 1957).

Digestibility of dry matter (Preston *et al.*, 1958) and particularly of crude fibre (Lassiter, Denton, Davis and Taylor, 1954, and Thompson and Grainger, 1954) and cellulose (Lodge, Miles and Jacobson, 1954 and others) is reduced by feeding antibiotics to ruminants. Ammonia concentration in the rumen is diminished (Preston *et al.*, 1958); there are lower levels of non-protein nitrogen in the blood (Lassiter, Denton and Rust, 1955) and this results in overall increased nitrogen retention (Hibbs, Conrad and Pounden, 1954; Preston *et al.*, 1958) in ruminants fed antibiotics.

Blood sugar levels are higher in calves fed chlortetracycline than in controls (Preston *et al.*, 1958).

Radisson, Smith and Ward (1956) studied calves fed whole milk + hay and grain to appetite. The influence of terramycin added to the milk was greatest during the first few weeks of life, and at this time the growth of the control (no terramycin) group of calves was depressed, suggesting that terramycin relieved a growth depression rather than stimulated growth. Oxytetracycline was found in the faeces 6–9 hours after first administration and high concentrations were found afterwards. Total and coliform counts did not change, but oxytetracycline lowered the numbers of enterococci, and also impaired their fermentative ability. This suggests that the presence of the enterococci limited the growth of the control calves. There was an increased resistance of the intestinal bacteria to antibiotics after antibiotic feeding, but antibiotic-resistant bacteria are said to have in general a lower metabolic rate and less virulence than sensitive bacteria (Pratt and Dufrenoy, 1949). In a later paper Radisson, Bartley, Lord and Swenson (1956) showed that aureomycin, fed in milk or in a capsule to calves receiving milk + hay + calf 'starter' ration, increased the sensitivity of intestinal bacteria (salmonella and coliform spp. were tested) to phagocytosis. They suggest that, although a change in numbers or types of intestinal bacteria is not always found on feeding antibiotics, these effects on the physiology of the potentially virulent bacteria could explain the effects of the antibiotics on growth, even when 'resistant' strains of bacteria develop. It might also explain why the greatest effect is found in the earlier weeks of life as the leucocytes from calves of 1–2 weeks old have a lower phagocytic activity than the cells from calves of 5–7 weeks old, and the younger calves are more susceptible to infections.

Lev and Briggs (1957) suggest that the increased growth of chicks on a diet containing penicillin is due to suppression of toxic bacteria in the intestine (in this case *Cl. welchii*) which decrease normal growth.

Since early-weaned calves, and those beef cattle for which growth responses have been reported, are fed rations composed mainly of concentrates, the effect of antibiotic feeding on these could be a mixture of the above effects together with suppression of rumen fermentation. For in the ruminant fed concentrates there is probably less metabolic loss when food is digested in the lower gut than when it is fermented in the rumen.

It is also considered by some that one effect of reduced fermentative activity in the rumen of animals fed antibiotics is a reduced gas production (both by fermentation and from bicarbonate in the saliva). There is evidence that gas pressure in the rumen controls appetite (as occurs in bloated animals), and if this is so the feeding of antibiotics, by reducing gas production in the rumen, should result in greater food consumption and greater growth. Dr. Johns and his colleagues in New Zealand have treated bloat by daily doses of penicillin during the critical period but a firm statement cannot yet be made regarding its efficacy.

While all this is an area of research which is being most actively prosecuted, the overall effects produced may have to be weighed against any untoward clinical conditions should they arise from a change in the microflora.

The Passage of Food through the Alimentary Canal

The various studies on the development of the ruminant stomach leave no doubt that the rumen, reticulum and omasum are of gastric origin.

The physical form of food and the amount given modify its passage through the alimentary canal, particularly through the rumen. The rate of passage of food by the suckling animal given milk is much more rapid than when solid food enters the rumen. There is undoubtedly a very large physical distension component regulating appetite (Blaxter, Graham and Wainman, 1956; Blaxter and Graham, 1956).

Probably the best of the various kinds of markers for studying the rate of passage of foodstuffs has been the stained straw particle technique of Lenkeit and Habeck (1930). Usuelli (1933) showed that marker-stained oats left the rumen in an exponential manner. Balch (1950) using this technique found that long hay was excreted more slowly than small particles of hay when these were fed together and ground hay was excreted more slowly than long hay when these were fed alone. The low digestibility that occurs on an all-ground ration cannot be explained on this basis, but particle size is of great importance. It is probable that the act of rumination by grinding larger particles into smaller ones speeds the passage of food through the rumen.

Since the rate of reticular contraction is fastest during eating (Schalk and Amadon 1928; Balch, 1952), there is reason for expecting maximum passage of digesta from the reticulo-rumen to the omasum during meals, and food leaving the reticulum and entering the omasum represents a mixture of several previous meals (Balch, Kelly and Heim, 1951).

Gray, Pilgrim and Weller (1953) from a study of the nitrogen, lignin and dry matter of the rumen and abomasal contents conclude that, in the sheep, from 33 to 64 per cent of the water entering the omasum is absorbed. They could find no evidence to suggest that a differential passage of fluid and solid occurred to any extent in the omasum. On the other hand, in the light of pressure changes in the omasum Balch *et al.* (1951) suggested that the digesta entering the neck of the omasum are pressed against the omasal leaves and that water and smaller particles of food carried with the water are more likely to enter the spaces between the leaves than larger particles. The view was expressed that most of the solid digesta pass through the neck without passing between the leaves. But the X-ray studies of Benzie and Phillipson (1957) seem to suggest rather the opposite.

The absorptive capacity of the omasum is recognised for fatty acids (Barcroft, McAnally and Phillipson, 1944) and for bicarbonate (including CO_2) (Ekman and Sperber, 1953), and there is little doubt that these anions are replaced by chloride from the blood (Ekman and Sperber, 1953); nitrogen is also absorbed (Badawy, Campbell, Cuthbertson and Mackie, 1958).

Balch (1957) found that on average 43 per cent of the dietary dry matter was apparently digested in the reticulo-rumen and the proportion of the diet was much higher with diets rich in concentrates than with those consisting mainly or entirely of roughage.

It would seem that when cows go out to graze in Spring the weight loss is mainly due to a reduction in the contents of the reticulo-rumen (Balch and Line, 1957).

The outflow of digesta from the abomasum appears to be continuous, although periods from 15 to 30 minutes occur when no flow is found (Phillipson, 1952), particularly with hay. Rumination causes a more regular flow. About 1·5 to 2 parts of gastric juice are needed to acidify 1 part of rumen or omasal contents.

Rumination

Rumination is a complex act. Above all, its biological value remains unknown. Insalivation, trituration, oxygen supply (Broberg, 1957) and rate of passage of food (Gordon, 1955) have all been suggested. Liquid is swallowed at intervals very soon

after the cud reaches the mouth on a wave of antiperistalsis, and while the animal is chewing. Finally the bolus of solid matter is swallowed and returns to the reticulo-rumen.

Rumination goes on mainly at night, particularly in the period 0400 to 0800 hours (Gordon, unpublished). The only condition that seems to alter the pattern of rumination is to grind all the food to a fine powder, which reduces the time devoted to this occupation. This is presumably due to lack of stimulation to the sensory pathways.

The Products of Reticulo-rumen Fermentation for Absorption

It is now firmly established that the volatile fatty acids are absorbed from the rumen into the blood stream. The rumen epithelium is also capable of metabolising each of the short-chain fatty acids produced by microbial decomposition of food in the rumen. The actual proportion of each fatty acid metabolised depends on the rate at which it is absorbed, which in turn depends on its concentration in the rumen. The mixture entering the blood stream from the rumen is similar to the mixture in the rumen but contains less butyrate because butyrate is metabolised by the epithelium and a considerable part is converted into ketone bodies (Pennington, 1952).

It is difficult to get an exact measure of the fatty acids produced and absorbed mainly from the rumen from any given ration. It has been calculated that for hay-fed and grain + hay-fed sheep it is about 70 per cent of the total energy requirements (Carroll and Hungate, 1954).

The low concentration of glucose in the blood of ruminants is well established. The quantity of glucose entering the portal blood is small compared with the concentrations of volatile fatty acids (Schambye, 1951*a*, *b*, *c*). The short-chain fatty acids are thus quantitatively much more important than glucose as a source of energy to the ruminant. Acetic acid accounts for most of the circulating lower fatty acid.

As to the metabolism of these acids, propionate is known to be a source of carbohydrate. Butyrate is more doubtful: it is generally acknowledged to be ketogenic. There is some evidence that it can cause a rise in the level of blood glucose, lactate and pyruvate in the sheep. Although acetic acid is quantitatively the most important and its significance is also most controversial, there is little doubt, as will be seen, that it contributes to the large heat increment that follows feeding and which is characteristic of the ruminant and an asset in cold or even temperate climates. Most tissues can oxidise it and it can be used by the mammary gland to synthesise some of the fatty acids in milk (McClymont, 1951; Popják, French, Hunter and Martin, 1951; and Cowie, Duncombe, Folley, French, Glascock, Massart, Peeters and Popják, 1951).

In diets low in hay and high in concentrates less acetate is produced by fermentation in the reticulo-rumen and low fat percentages are found in the milk due to the shortage of acetate as a raw material for fat synthesis. It has been shown that milk fat can be increased by the addition of sodium acetate to the diet (National Institute for Research in Dairying, 1956). Ammonia is constantly present and its concentration varies to a great extent with the nature and quantity of protein in the feed and with the quantity of starch or other soluble carbohydrate in the feed.

In all probability much of the B vitamin content of the rumen is in true solution —except possibly B_{12} and its associated components. Calcium, magnesium, sodium, potassium, chloride and HCO_3 ions are all present together with the trace elements in the rumen contents. Sodium can move from the reticulo-rumen contents against the electrochemical p.d. of the ion into the plasma. The activity of this sodium 'pump' is important to the sodium economy of the animal (Dobson, 1956).

The gases are principally carbon dioxide and methane, while oxygen and nitrogen

have been found in small quantities that increase as the time after feeding becomes prolonged (Washburn and Brody, 1937). The carbon dioxide is derived from both the food and the saliva and forms more than 60 per cent of the gas present. Methane is formed as a result of the reduction of carbon dioxide by rumen bacteria. Five to 12 per cent of the energy of the ration is lost as methane in digestion (Reid, 1956).

It seems reasonably certain from the work of Dr. K. L. Blaxter and his colleagues at the Hannah Institute that in animals given food sufficient to meet their requirement for maintenance, the whole of the carbohydrate dissimilated could be absorbed as fatty acids and yet the heat increment would be low. It would be higher than that observed in man given a comparable amount of digestible carbohydrate, since the heat increment of the acids is 15 per cent in the sheep, whereas in man the heat increment of hexose is about 8 per cent (Armstrong and Blaxter, 1957; Armstrong, Blaxter and Graham, 1957).

The Abomasum

The digesta entering the abomasum have lost about 70 per cent of the digestible fibre present, most of the starch and pentosans, and the soluble sugars have also been fermented. It is not known how much protein and non-protein nitrogen is converted to microbial substance. Experimentally it was found that 90 per cent of casein in a partially purified diet for sheep was found to be converted into microbial protein (McDonald and Hall, 1957). According to Hogan's work (1957) some 36 per cent of the digestible nitrogen is absorbed in the stomach and about 6½ per cent in the small intestine.

The gastric juice which is similar to that of other non-ruminant farm animals is secreted in quantities that give a dilution of about 1:1 to 2:1 for rumen contents and there the non-protein nitrogen fraction is increased owing to peptic digestion. Protozoa are digested and the bacteria lose their definition and staining properties.

Heald (1951) estimated that in the hay-fed sheep about 5 to 6 g. of microbial polysaccharide reached the abomasum. Estimates of glucose absorbed into the portal stream when compared to volatile fatty acids in sheep on various rations supports the concept that glucose derived by the animal from microbial synthesis of polysaccharides is comparatively small.

From this stage onwards the course of digestion is believed to resemble that of the non-ruminant.

Summary

The pre-gastric microbial digestion of food confers important benefits upon the ruminant. Energy from fibrous foods, which would not otherwise be obtained, is more efficiently garnered and the available energy is yielded over a longer period of time so that the animal is more resistant to periods of starvation; the uptake of nitrogen is controlled and that from non-protein sources utilised more efficiently via microbial proteins; and considerable synthesis of vitamins takes place. This aspect of the physiology of digestion is as fascinating for the investigator as it is important for the husbandman. Further advances should help to clear our understanding of some of the so-called 'metabolic' diseases.

Bibliography

Annison, E. F. (1954): *Biochem. J.* **57,** 400.
Armstrong, D. G. & Blaxter, K. L. (1957): *Brit. J. Nutrit.* **11,** 247.
Armstrong, D. G., Blaxter, K. L. & Graham, N. McC. (1957): *Brit. J. Nutrit.* **11,** 392.
Badawy, A. M., Campbell, R. M., Cuthbertson, D. P. & Mackie, W. S. (1958): *Brit. J. Nutrit.* In press.
Baker, F. (1946): *Nature, Lond.* **158,** 609.
Balch, C. C. (1950): *Brit. J. Nutrit.* **4,** 361.
Balch, C. C. (1952): *Brit. J. Nutrit.* **6,** 366.
Balch, C. C. (1957): *Brit. J. Nutrit.* **11,** 213.

Balch, C. C., Kelly, A. & Heim, G. (1951): *Brit. J. Nutrit.* **5,** 207.

Balch, C. C. & Line, C. (1957): *J. Dairy Res.* **24,** 11.

Barcroft, J., McAnally, R. A. & Phillipson, A. T. (1944): *J. exp. Biol.* **20,** 120.

Becker, E. R. & Hsiung, T. S. (1929): *Proc. nat. Acad. Sci.* **15,** 684.

Benzie, D. & Phillipson, A. T. (1957): *The Alimentary Tract of the Ruminant.* Oliver & Boyd, Edinburgh and London.

Blaxter, K. L. & Graham, N. McC. (1956): *J. agric. Sci.* **47,** 207.

Blaxter, K. L., Graham, N. McC. & Wainman, F. W. (1956): *Brit. J. Nutrit.* **10,** 69.

Blaxter, K. L. & Wood, W. A. (1952): *Brit. J. Nutrit.* **6,** 56.

Bouchaert, J. H. & Oyaert, W. (1952): *Zootechnia,* **1,** 21.

Broberg, G. (1957): *Nord. VetMed.* **2,** 57.

Burroughs, W., Latona, A., De Paul, P., Gerlaugh, P. & Bethke, R. M. (1951): *J. Anim. Sci.* **10,** 693.

Carroll, E. J. & Hungate, R. E. (1954): *App. Microbiol.* **2,** 205.

Chalmers, M. I. & Synge, R. L. M. (1954*a*): *J. agric. Sci.* **44,** 263.

Chalmers, M. I. & Synge, R. L. M. (1954*b*): *Advanc. in Protein Chem.* **9,** 93.

Chalmers, M. I. & Synge, R. L. M. (1957): *Collected Papers of Rowett Research Institute,* p. 37.

Chanda, R. & Owen, E. C. (1952): *Biochem. J.* **51,** 100.

Colin, M. G. (1886): *Traité de Physiologie Comparée des Animaux.* Baillière et Fils, Paris.

Cowie, A. T., Duncombe, W. G., Folley, S. J., French, T. H., Glascock, R. F., Massart, L., Peeters, G. J. & Popják, G. (1951): *Biochem. J.* **49,** 610.

Cunningham, H. M. & Loosli, J. K. (1954): *J. Anim. Sci.* **13,** 265.

Cuthbertson, D. P. (1953): *Advancement of Science,* **10,** 202.

Denton, D. A. (1957): *Quart. J. exp. Physiol.* **42,** 72.

Dobson, A. (1956): *The movement of ions across the epithelium of the reticulo-rumen sac.* Ph.D. Thesis. University of Aberdeen.

Doetsch, R. N., Howard, B. H., Mann, S. O. & Oxford, A. E. (1957): *J. gen. Microbiol.* **16,** 156.

Doetsch, R. & Robinson, R. Q. (1953): *J. Dairy Sci.* **36,** 115.

Dollar, A. M. (1958): Ph.D. Thesis, University of Reading.

Dollar, A. M. & Porter, J. W. G. (1957): *Nature,* Lond. **179,** 1299.

Duncan, C. W., Agrawala, I. P., Huffman, C. F. & Luecke, R. W. (1953): *J. Nutrit.* **49,** 41.

Eadie, J. M. (1957): *Proc. roy. Soc. Edinb.* B. **66,** 276.

Eadie, J. M. & Oxford, A. E. (1957): *Nature,* Lond. **179,** 485.

Edwards, D. C. (1955): *J. Dairy Res.* **22,** 232.

Elsden, S. R., Volcani, B. E., Gilchrist, F. M. C. & Lewis, D. (1956): *J. Bact.* **72,** 681.

Ekman, J. & Sperber, I. (1953): *Ann. agr. Coll. Sweden,* **19,** 227.

Ford, J. E., Kon, S. K. & Porter, J. W. G. (1951): *Biochem. J.* **50,** ix.

Garton, G. A. & Oxford, A. E. (1955): *J. Sci. Fd Agric.* **6,** 142.

Gordon, J. G. (1955): Ph.D. Thesis, University of Aberdeen.

Gray, F. V. (1947): *J. exp. Biol.* **24,** 1.

Gray, F. V., Pilgrim, A. F., Rodda, H. J. & Weller, R. A. (1951): *Nature,* Lond. **167,** 954.

Gray, F. V., Pilgrim, A. F., Rodda, H. J. & Weller, R. A. (1952): *J. exp. Biol.* **29,** 57.

Gray, F. V., Pilgrim, A. F. & Weller, R. A. (1953): *Nature,* Lond. **172,** 347.

Gutierrez (1953): *J. Bact.* **66,** 123.

Heald, P. J. (1951): *Brit. J. Nutrit.* **5,** 84.

Henneberg, W. & Stohmann, F. (1885): *Hoppe-Seyler's Z.* **21,** 613.

Hibbs, J. W., Conrad, H. R. & Pounden, W. D. (1954): *J. Dairy Sci.* **37,** 724.

Hill, K. J. (1955): *Quart. J. exp. Physiol.* **40,** 32.

Hobson, P. N. & Mann, S. O. (1955): *J. gen. Microbiol.* **13,** 420.

Hobson, P. N., Mann, S. O. & Oxford, A. E. (1958): *J. gen. Microbiol.* **18,** and to be published.

Hogan, J. P. (1957): *J. Physiol.* **139,** 25*P*.

Howard, B. H. (1955): *Biochem. J.* **60,** i.

Howard, B. H. (1957): *Biochem. J.* **67,** 18*P*.

Huffman, C. F. (1953): *Annu. Rev. Biochem.* **22,** 399.

Huhtanen, C. N., Saunders, R. K. & Gall, L. S. (1951): *J. Anim. Sci.* **10,** 1049.

Hungate, R. E. (1950*a*): *Bact. Rev.* **14,** 1.

Hungate, R. E. (1950*b*): *Ann Rev. Microbiol.* **4,** 53.

Johns, A. T. (1951): *J. gen. Microbiol.* **5,** 317.

Johns, A. T., Mangan, J. L. & Reid, C. S. W. (1957): *N.Z. Vet. J.* **5,** 115.

Kellner, O. (1913): *The scientific feeding of animals.* Translated by W. Goodwin; Duckworth, London.

Kon, S. K. (1955): *Symp. biochem. Soc. No. 13,* 17.

Kon, S. K. & Henry, K. M. (1949): *J. Dairy Res.* **16,** 68.

Kon, S. K. & Henry, K. M. (1951): *J. Dairy Res.* **18,** 317.

Kon, S. K. & Porter, J. W. G. (1954): In *Vitamins and Hormones. Advances in Research and Application.* Ed. by R. S. Harris, G. F. Marrian & K. V. Thimann. Vol. XII, pp. 53–65. Acad. Press Inc., New York.

Lassiter, C. A. (1955): *J. Dairy Sci.* **38,** 1102.

Lassiter, C. A., Denton, T. W., Davis, C. L. & Taylor, J. A. (1954): *J. Anim. Sci.* **13,** 991.

Lassiter, C. A., Denton, T. W. & Rust, J. W. (1955): *J. Anim. Sci.* **14,** 760.

Lengemann, F. W. & Allen, N. N. (1955): *J. Dairy Sci.* **38,** 651.

Lenkeit, W. & Habeck, R. (1930): *Arch. Tierernähr. Tierz.* **2,** 517.

Ley, M. & Briggs, C. A. E. (1957): *Brit. J. Nutrit.* **11,** 364.

Lodge, J. R., Miles, J. T. & Jacobson, N. L. (1954): *J. Dairy Sci.* **37,** 663.

LOOSLI, J. K., WILLIAMS, H. H., THOMAS, W. E., FERRIS, F. H. & MAYNARD, L. A. (1949): *Scien* **110,** 144.
MANN, S. O., MASSON, F. M. & OXFORD, A. E. (1954): *Brit. J. Nutrit.* **8,** 246.
MANN, S. O. & OXFORD, A. E. (1954): *J. gen. Microbiol.* **11,** 83.
MANN, S. O. & OXFORD, A. E. (1955): *J. gen. Microbiol.* **12,** 140.
MARSTON, H. R. (1948): *Biochem. J.* **42,** 564.
MARSTON, H. R. & SMITH, R. M. (1952): *Nature, Lond.* **170,** 792.
MAYNARD, L. A. (1957): *Nutrition and Digestive Flora in Animals.* Fourth Internat. Cong. Nutritio Paris. P. 179.
MCCLYMONT, G. L. (1951): *Austral. J. agric. Res.* **2,** 158.
MCDONALD, I. W. (1948): *Biochem. J.* **42,** 584.
MCDONALD, I. W. & HALL, R. J. (1957): *Biochem. J.* **67,** 400.
MCNAUGHT, M. L., OWEN, E. C., HENRY, K. M. & KON, S. K. (1954): *Biochem. J.* **56,** 151.
MCNAUGHT, M. L., OWEN, E. C. & SMITH, J. A. B. (1950): *Biochem. J.* **46,** 36.
MILLS, C. F. (1958): *Soil Science,* **85,** 100.
MOIR, R. J., SOMERS, M. & WARING, H. (1956): *Austral. J. biol. Sci.* **9,** 293.
MUNRO, J. (1956): *Brit. J. anim. Behaviour,* **4,** 34.
NATIONAL INSTITUTE FOR RESEARCH IN DAIRYING (1956): *Ann. Rep.*
OWEN, E. C. (1954): *J. Dairy Res.* **21,** 408.
OXFORD, A. E. (1951): *J. gen. Microbiol.* **5,** 83.
PEARSON, R. M. & SMITH, J. A. B. (1943): *Biochem. J.* **37,** 153.
PENNINGTON, R. J. (1952): *Biochem. J.* **51,** 25.
PFIZER & CO. (1957): *Tech. Bull. Series* 26.
PHILLIPSON, A. T. (1942): *J. exp. Biol.* **19,** 186.
PHILLIPSON, A. T. (1952): *J. Physiol.* **116,** 84.
PHILLIPSON, A. T. & CUTHBERTSON, D. P. (1956): *Modern Concepts of Rumen Digestion and Me bolism.* Seventh Internat. Cong. Anim. Husbandry, Madrid. P. 1.
PIGDEN, W. J. & STONE, J. E. (1952): *Sci. Agric.* **32,** 502.
POPJÁK, G., FRENCH, T. H., HUNTER, G. D. & MARTIN, A. J. P. (1951): *Biochem. J.* **48,** 612.
PORTER, J. W. G. (1953): *Proc. Nutr. Soc.* **12,** 106.
PORTER, J. W. G. (1955): *National Institute for Research in Dairying Annual Rep.,* p. 101.
POUNDEN, W. D. & HIBBS, J. W. (1948*a*): *J. Dairy Sci.* **31,** 1041.
POUNDEN, W. D. & HIBBS, J. W. (1948*b*): *J. Dairy Sci.* **31,** 1051.
POUNDEN, W. D. & HIBBS, J. W. (1949): *J. Dairy Sci.* **32,** 1025.
POUNDEN, W. D. & HIBBS, J. W. (1950*a*): *J. Dairy Sci.* **33,** 378.
POUNDEN, W. D. & HIBBS, J. W. (1950*b*): *J. Dairy Sci.* **33,** 639.
PRATT & DUFRENOY (1949): *Antibiotics.* Lippincott Co., Philadelphia. *loc. cit.* Radisson, Smith Ward (1956).
PRESTON, T. R. (1956): *Agriculture, Lond.* **62,** 464.
PRESTON, T. R., MCLEOD, N. A. & DINDA, P. K. (1958): *J. Anim. Prod.* In press.
RADISSON, J. J., BARTLEY, E. E., LORD, T. H. & SWENSON, M. J. (1956): *J. Dairy Sci.* **39,** 1386
RADISSON, J. J., SMITH, C. K. & WARD, G. M. (1956): *J. Dairy Sci.* **39,** 1260.
REID, J. T. (1953): *J. Dairy Sci.* **36,** 955.
REID, J. T. (1956): *Cornell Agr. Expt. Sta. Mem.* 344.
REISER, R. (1951): *Fed. Proc.* **10,** 236.
SCHALK, A. F. & AMADON, R. S. (1928): *Bull. N. Dak. Agric. Exp. Sta. No. 216.*
SCHAMBYE, P. (1951*a*): *Nord. VetMed.* **3,** 555.
SCHAMBYE, P. (1951*b*): *Nord. VetMed.* **3,** 748.
SCHAMBYE, P. (1951*c*): *Nord. VetMed.* **3,** 1003.
SCHEUNERT, A. & TRAUTMANN, A. (1921): *Pflüg. Arch.* **192,** 33.
EL-SHAZLY, K. (1952*a*): *Biochem. J.* **51,** 640.
EL-SHAZLY, K. (1952*b*): *Biochem. J.* **51,** 647.
SHORLAND, F. B., WEENICK, R. O. & JOHNS, A. T. (1955): *Nature,* Lond. **175,** 1129.
TAPPEINER, H. (1884): *Z. Biol.* **20,** 52.
THOMPSON, C. M. & GRAINGER, R. B. (1954): *J. Anim. Sci.* **13,** 1002.
UNDERWOOD, E. J. & MOIR, R. J. (1953): *J. Aust. Inst. agric. Sci.* **19,** 214.
USUELLI, F. (1933): *Profilassi,* **6,** 7.
USUELLI, F. (1949): *Zoot. e Vet.* **4,** 3.
USUELLI, F. & FIORINI, P. (1938): *Boll. Soc. ital. Biol. oper.* **13,** 4.
WARNER, A. C. I. (1956): *Biochem. J.* **64,** 1.
WARNER, A. C. I. (1956): *J. gen. Microbiol.* **14,** 733.
WASHBURN, L. E. & BRODY, S. (1937): *Mo. Agric. Expt. Sta. Res. Bull.* 263.
WASSERMAN, R. H., SEELEY, H. W. & LOOSLI, J. K. (1953): *J. Anim. Sci.* **12,** 935.
WELLER, R. A. & GRAY, F. V. (1954): *J. exp. Biol.* **31,** 40.